Statistical Analysis Plan J2A-JE-GZPD (2.0)

A Phase 3, Randomized, Double-Blind Study to Investigate the Efficacy and Safety of Once-Daily Oral LY3502970 Compared with Placebo in Japanese Adult Participants with Obesity Disease (ATTAIN-J)

NCT05931380

Approval Date: 05-Aug-2025

Title Page

Protocol Title: A Phase 3, Randomized, Double-Blind Study to Investigate the Efficacy and Safety of Once-Daily Oral LY3502970 Compared with Placebo in Japanese Adult Participants with Obesity Disease (ATTAIN-J)

Protocol Number: J2A-JE-GZPD

Compound Number: Orforglipron (LY3502970)

Short Title: GZPD SAP

Acronym: ATTAIN-J

Sponsor Name: Eli Lilly Japan K.K.

Legal Registered Address: LILLY PLAZA ONE BLDG., 5-1-28 Isogamidori, Chuo-ku Kobe City, Hyogo 651-0086 Japan

Regulatory Agency Identifier Number(s): Not applicable

Document ID: VV-CLIN-137368

Table of Contents

Title Page 1
Table of Contents 2
Version history 6
1. Introduction 9
1.1. Objectives, Endpoints, and Estimands 10
1.2. Study Design 18
2. Statistical Hypotheses 20
2.1. Multiplicity Adjustment 21
3. Analysis Sets 22
4. Statistical Analyses 24
4.1. General Considerations 24
4.1.1. Definition of Baseline 25
4.1.2. Analysis Methods 26
4.2. Participant Dispositions 35
4.3. Primary Endpoints/Estimands Analysis 36
4.3.1. Definition of Endpoint(s) 36
4.3.2. Main Analytical Approach 36
4.3.3. Sensitivity Analyses 36
4.3.4. Supplementary Analyses 36
4.4. Secondary Endpoints/Estimand Analysis 37
4.4.1. Secondary Endpoints 61
4.4.2. Additional Secondary Endpoints 61
4.5. Exploratory Endpoints Analysis 61
4.6. Safety Analyses 61
4.6.1. Extent of Exposure 62
4.6.2. Adverse Events 62
4.6.3. Clinical Laboratory Evaluation 63
4.6.4. Vital Signs and Physical Characteristics 63
4.6.5. Electrocardiograms 64
4.6.6. Safety Topics of Interest 64
4.7. Other Analyses 71
4.7.1. Subgroup Analyses 71
4.8. Interim Analyses 72
4.8.1. Data Monitoring Committee 72
4.9. Changes to Protocol-Planned Analyses 72
5. Sample Size Determination 73

Approved on 05 Aug 2025 GMT

6. Supporting Documentation....74
6.1. Appendix 1: Demographic and Baseline Characteristics....74
6.2. Appendix 2: Historical Illnesses and Preexisting Conditions....76
6.3. Appendix 3: Concomitant Medications....77
6.4. Appendix 4: Treatment Compliance....79
6.5. Appendix 5: Important Protocol Deviations....79
6.6. Appendix 6: Clinical Trial Registry Analyses....80
7. References....81

Table of Contents

Tables **Page**

Table GZPD.3.1. Participant Analysis Sets 22
Table GZPD.3.2. Data Points Sets for Non-Glycemic-Related Parameters 23
Table GZPD.3.3. Data Points Sets for Glycemic-Related Parameters 23
Table GZPD.3.4. Data Points Sets for Safety Analyses 23
Table GZPD.4.1. Baseline and Postbaseline Definitions for Safety Analyses 26
Table GZPD.4.2. Imputation Procedure 29
Table GZPD.4.3. Categorization of Treatment Period Discontinuation Reasons 30
Table GZPD.4.4. Listings and Summary Tables Related to Dispositions 36
Table GZPD.4.5. Description and Derivation of Primary and Secondary Endpoints for Efficacy Measures and Patient Reported Outcomes 38
Table GZPD.4.6. Description of Efficacy/Health Outcomes Analyses 52
Table GZPD.4.7. Summary Tables Related to Adverse Events 63
Table GZPD.4.8. Summary Tables Related to Clinical Laboratory Evaluations. 63
Table GZPD.4.9. Summary Tables Related to Vital Signs 64
Table GZPD.4.10. Summary Tables Related to ECG Parameters 64
Table GZPD.4.11. Description and Analyses of Safety Topics of Interest 65
Table GZPD.6.1. Demographics and Baseline Characteristics with Variables for Subgroup Analysis 75
Table GZPD.6.2. Summary Tables Related to Historical Illness and Preexisting Conditions 77
Table GZPD.6.3. Summary Tables Related to Concomitant Medications 77
Table GZPD.6.4. Protocol-Specified Concomitant Medications 78
Table GZPD.6.5. Prohibited Concomitant Medication Rules 78

Approved on 05 Aug 2025 GMT

Table of Contents

Figure **Page**

Figure GZPD.1.1. Illustration of study design for Clinical Protocol J2A-JE-GZPD. 19

Figure GZPD.2.1. Graphical testing procedure. 21

Figure GZPD.4.1. Treatment journey of a trial participant.....27

Version history

This Statistical Analysis Plan (SAP) Version 1 for Study J2A-JE-GZPD (GZPD) is based on GZPD Protocol Amendment (b) dated 7 December 2023.

The major updates in this SAP Version 2 are listed in the table below.

SAP Version History Summary

SAP Version	Approval Date	Section	Change	Rationale
1	September 20, 2024		• Not Applicable	• Original version
2	See date on signature page	2.1	• Updated graphical testing scheme with final version	• Clarification
		3	• Removed note regarding no inferential comparisons for the summary of disposition, demographics, historical illness, and preexisting conditions	• Clarification
		3	• Clarified only 1 safety data points set	• Clarification
		4.1	• Updated strata variable definition	• To consider the small size of strata
		4.1.2.1.1	• Updated analysis model	• According to strata variable change
		4.1.2.1.2	• Further clarified the definition of Scenario 5. • Removed the direct	• The direct imputation method is not fully aligned with the primary imputation strategy.

			imputation method • Updated the imputation method for Scenario 5B • Added footnote in Table 4.3 for emergency unblinding	• Used a different method to handle Scenario 5B, based on simulation results from Liu et al. (2024) • Clarification
		4.1.2.1.3	• Removed the penalty range for the MI-TP analysis to -15% to 15%	• Updated because of data dependency
		4.1.2.2.3	• Added a censoring condition for date of initiation of prohibited medications	• To align with the defined estimand
		4.1.2.3	• Updated MMRM model for safety analysis • Clarified the definition of start time for different analyses	• To be consistent with the PSAP
		Table GZPD.4.5	• Updated eGFR estimation method to using JSN eGFR	• To be consistent with the protocol
		4.6.6	• Modified Table GZPD.4.11	• To be consistent with the PSAP
		4.7.1	• Updated the testing of subgroup effects.	• Clarification

Approved on 05 Aug 2025 GMT

			• Updated the Bayesian shrinkage method	
		Table GZPD.6.1	• Updated the BMI category. • Updated eGFR estimation method. • Define obesity related health disorders	• Clarification

1. Introduction

This SAP is intended to describe the analyses of primary and secondary objectives, as well as safety assessments, for Study GZPD. Sensitivity and supplementary analyses intended to support the primary objectives are also included. Additional exploratory analyses, if needed, may be included in a separate document.

Pharmacokinetic/pharmacodynamic (PK/PD) analyses will be documented in a separate document.

Changes to the protocol-planned analyses are described in Section 4.9.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To demonstrate that orforglipron 6 mg, 12 mg, and 36 mg QD is superior to placebo in change from baseline for: • body weight.	From baseline to Week 72 • mean percent change in body weight, and • percentage of participants who achieve ≥5% body weight reduction.
Secondary	
To demonstrate that orforglipron 6 mg, 12 mg, and 36 mg QD is superior to placebo in change from baseline for the following: • body weight. • Improvement of comorbidities	From baseline to Week 72 • Percentage of participants who achieve: o ≥7% body weight reduction o ≥10% body weight reduction o ≥15% body weight reduction, and o ≥20% body weight reduction. • Mean change in: o body weight (kg), and o BMI (kg/m^2). • Percentage of participants who had improvements in: o hypertension (only for participants with hypertension at baseline) o dyslipidemia (only for participants with dyslipidemia at baseline), and o T2D (<6.5% [48 mmol/mol] in HbA1c) (only for participants with T2D at baseline). • Percentage of participants who achieve HbA1c target value at Week 72 (only for participants with T2D at baseline): o <7.0% (53 mmol/mol), and o <5.7% (39 mmol/mol). • Change in anti-hypertensive medication (decrease, no change, increase) (only hypertension at baseline) • Change in lipid-lowering medication (decrease, no change, increase) (only for participants with dyslipidemia at baseline) • Change in concomitant oral anti-glycemic medication (decrease, no change, increase) (only for participants with T2D at baseline), and

Objectives	Endpoints
	• Change in glycemic category (normo-glycemia, borderline type-diabetes, T2D).
• VAT and SAT in CT subgroups	• Proportion of participants who achieved the following number of improvements of obesity-related health problems (hypertension, dyslipidemia, and T2D) at Week 72 from baseline, regardless of BMI at baseline o achieved 0, 1, 2 or 3 improvements (only for participants with 3 obesity-related health problems) o achieved 0, 1 or 2 improvements (only for participants with 2 obesity-related health problems), and o achieved 0 or 1 improvements (only for participants with 1 obesity-related health problems). • Mean change in o VAT (cm^2) o SAT (cm^2), and o VAT/SAT ratio • Percentage of participants who achieve VAT <100 cm^2 (only for participants with VAT ≥100 cm^2 at baseline).
To demonstrate that orforglipron 6 mg, 12 mg, and 36 mg QD is superior to placebo in change from baseline for the following: • Waist circumference • Blood pressure in overall and hypertension at baseline • Lipid in overall and dyslipidemia at baseline	From baseline to Week 72 • Mean change in waist circumference at umbilical level (cm) • Mean change in: o systolic BP (mm Hg) o diastolic BP (mm Hg) • Mean change in: o total cholesterol (mg/dL) o non-HDL cholesterol (mg/dL) o triglyceride (mg/dL) o LDL cholesterol (mg/dL) o HDL cholesterol (mg/dL) o VLDL cholesterol (mg/dL), and o free fatty acid (mmol/L, mEq/L).

Approved on 05 Aug 2025 GMT

Objectives	Endpoints
• Glycemic control in overall and T2D at baseline	• Mean change in: o fasting blood glucose (mg/dL) o HbA1c (%, mmol/mol) o fasting insulin (mIU/L, pmol/L), and o fasting C-peptide (µg/L, nmol/L). • Mean change in (only T2D at baseline) o 7-point SMBG
• Biomarkers	• Mean change in: o hsCRP (mg/L) o Uric acid (mg/dL) o UACR (g/mol, g/kg) o eGFR (mL/min/1.73 m2)
• Patient reported outcomes	• Mean change in SF-36v2 acute form domain scores • Mean change in EQ-5D-5L health state utilities and VAS • Mean change in IWQOL-Lite-CT Physical Function, Physical, and Psychosocial composite score, and • Mean change in PGI-S and PGI-C limitations on physical function due to weight score.
To demonstrate orforglipron (6 mg, 12 mg, and 36 mg QD pooled dose) is superior to placebo for the following:	From baseline to Week 72
• Body weight	• Percentage of participants who achieve: o BMI <25 kg/m^2 and o BMI <27 kg/m^2.
• Improvement of comorbidities	• Percentage of participants who had improvements in: o hypertension (only for participants with hypertension at baseline) o dyslipidemia (only for participants with dyslipidemia at baseline) o T2D (<6.5% [48 mmol/mol] in HbA1c) (only for participants with T2D at baseline) o change in anti-hypertensive medication (decrease, no change, increase) (only for participants with hypertension at baseline) o change in lipid-lowering medication (decrease, no change, increase) (only for participants with dyslipidemia at baseline)

Objectives	Endpoints
	○ change in concomitant oral anti-glycemic medication (decrease, no change, increase) (only for participants with T2D at baseline), and ○ change in glycemic category (normo-glycemia, borderline type-diabetes, T2D).
	• Proportion of participants who achieved the following number of improvements of obesity-related health problems (hypertension, dyslipidemia, and T2D) at Week 72 from baseline, regardless of BMI at baseline ○ achieved 0, 1, 2 or 3 improvements (only for participants with 3 obesity-related health problems) ○ achieved 0, 1 or 2 improvements (only for participants with 2 obesity-related health problems) ○ achieved 0 or 1 improvements (only for participants with 1 obesity-related health problems)
To demonstrate orforglipron (6 mg, 12 mg, and 36 mg QD pooled dose) is superior to placebo in change from baseline for the following: • Biomarkers • Glycemic control (only for participants with T2D at baseline) • Delayed progression to T2D • Patient-reported outcomes	From baseline to Week 72 • Mean change in: ○ adiponectin (both high molecular weight (mg/L) and total adiponectin (µg/L)) ○ leptin (ng/L) ○ UACR (g/mol, g/kg) ○ eGFR (mL/min/1.73 m^2) • Mean change in: ○ HOMA2-IR ○ HOMA2-B • Time to onset of T2D (only nonT2D at baseline) • Mean change in SF-36v2 acute form scores: ○ physical functioning ○ role-physical score

Objectives	Endpoints
	○ bodily pain score ○ general health score ○ vitality score ○ social functioning score ○ role-emotional score ○ mental health score ○ physical component summary ○ mental component summary • Mean change in IWQOL-Lite-CT scores: ○ physical function score ○ physical score ○ psychosocial score ○ total score • Mean change in PGI-C and PGI-S scores: ○ Shift of the ordinal response category • Mean change in PROMIS Short Form Sleep Disturbance 8b score from baseline
To characterize the population PK of orforglipron	• Population PK parameters
Exploratory	
To explore the relationships between orforglipron concentration and efficacy, safety, and tolerability measures	• Population PD parameters

Abbreviations: BMI = body mass index; BP = blood pressure; CKD-EPI = Chronic Kidney Disease Epidemiology; CT = computerized tomography; eGFR = estimated glomerular filtration rate (calculated by Japanese Society of Nephrology; HbA1c = hemoglobin A1c; HDL = high density lipoprotein cholesterol; HOMA2-B = homeostasis model assessment estimates steady state beta cell function; HOMA2-IR = homeostasis model assessment estimates insulin resistance beta cell function; hsCRP = high-sensitivity C-reactive protein; IWQOL-Lite CT = Impact of Weight on Quality of Life-Lite Clinical Trials version; LDL = low density lipoprotein; PD = pharmacodynamics; PGI-C = Patient Global Impression of Change; PGI-S = Patient Global Impression of Status; PK = pharmacokinetics; PROMIS = Patient-Reported Outcomes Measurement Information System; QD = once daily; SF-36v2 = Short-Form 36-item Health Survey Version 2; SMBG = self-monitored blood glucose; T2D = type 2 diabetes; SAT = subcutaneous adipose tissue; UACR = urine albumin/creatinine ratio; VAT = visceral adipose tissue; VAS = visual analogue scale; VLDL = very low density lipoprotein.

Estimand

There will be 2 estimands planned in the study. These estimands address intercurrent events (ICEs) using either the treatment policy strategy or the hypothetical strategy (ICH 2019), respectively.

Estimands for weight-related, blood pressure, and lipid parameters

The below estimands will be applied to weight-related, blood pressure (BP) and lipid parameters, using percent change in body weight at Week 72 visit as an example. For other weight-related, BP, and lipid parameters, replace percent change in body weight by the endpoints to be analyzed.

Treatment regimen estimand

The treatment regimen estimand will be the primary estimand.

The clinical question of interest: *What is the treatment difference in the percent change in body weight and percentage of participants who achieve 5% or more body weight reduction from baseline to 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals who meet eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments?*

Treatment policy strategy

The occurrence of the ICE is considered irrelevant in defining the treatment effect of interest. As such, the values for the variable of interest are used regardless of whether the ICE occurs.

The treatment regimen estimand is described by the following attributes:

Population: individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH 2019). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.

Endpoints: percent change in body weight from baseline and percentage of participants who achieve 5% or more body weight reduction from baseline to 72 weeks.

Treatment condition: the randomized treatment with allowance for potential dose interruptions and modifications, regardless of adherence to study intervention or initiation of prohibited weight management treatments.

ICEs: no ICEs are defined because treatment adherence and the initiation of prohibited weight management treatments are a part of the treatment condition.

Population-level summary and treatment effect of interest: the difference in mean percent change in body weight and percentage of participants who achieve 5% or more body weight reduction from baseline to 72 weeks between orforglipron and placebo.

Rationale for the estimand: the estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population.

Efficacy estimand

The clinical question of interest: *What is the treatment difference in the percent change in body weight and percentage of participants who achieve 5% or more body weight reduction from baseline to 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals who meet the eligibility criteria if they would remain on their randomly assigned treatment for 72 weeks and would not initiate prohibited weight management treatments?*

Hypothetical strategy

A scenario is envisaged in which the ICE would not occur. The value of the variable to reflect the clinical question of interest is the value that the variable would have taken in the hypothetical scenario defined.

The efficacy estimand is described by the following attributes:

Population: individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.

Endpoint: percent change in body weight and percentage of participants who achieve 5% or more body weight reduction from baseline to 72 weeks.

Treatment condition: the randomized treatment with allowance for potential dose interruptions and modifications.

ICEs: ICEs include permanent discontinuation of study intervention and initiation of prohibited weight management treatments, which is handled by the hypothetical strategy. The potential outcome of interest is the response in the efficacy measurement if participants would remain on their randomly assigned study treatment for 72 weeks and would not initiate prohibited weight management treatments. Dose modification and interruption will not be considered as ICEs because they are part of the treatment condition.

Population-level summary and treatment effect of interest: difference in the mean percent change in body weight and percentage of participants who achieve 5% or more body weight reduction from baseline to 72 weeks between orforglipron and placebo.

Rationale for the estimand: this estimand aims to study the efficacy of orforglipron under the ideal condition that all participants adhere to their randomly assigned treatment without being confounded by the initiation of other weight management treatments.

Estimands for glycemic parameters

The below estimands will be applied to glycemic parameters, using change in hemoglobin A1c (HbA1c) at Week 72 as an example. For other glycemic parameters, replace change in HbA1c by the endpoints to be analyzed.

Treatment regimen estimand

The treatment regimen estimand will be the primary estimand.

The clinical question of interest: *What is the treatment difference in the change in HbA1c from baseline at 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals who meet the eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments or glycemic rescue therapy?*

The “treatment regimen” estimand is described by the following attributes:

- ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
- ***Endpoints:*** Change in HbA1c from baseline to 72 weeks.
- ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications regardless of adherence to study intervention or initiation of prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy.
- ***ICEs:*** No ICEs are defined because treatment adherence and the initiation of prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy are a part of the treatment condition.
- ***Population-level summary and treatment effect of interest:*** The difference in mean changes in HbA1c from baseline to Week 72 between orforglipron and placebo.

Rationale for the estimand: The estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population.

Efficacy estimand

The clinical question of interest: *What is the treatment difference in the change in HbA1c from baseline at 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg versus. placebo, as an adjunct to healthy diet and physical activity, in individuals who meet the eligibility criteria if they would remain on their randomly assigned treatment for 72 weeks and would not initiate prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy?*

The "efficacy" estimand is described by the following attributes:

- ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
- ***Endpoint:*** Change in HbA1c from baseline to 72 weeks.
- ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications.
- ***ICEs:*** ICEs include permanent discontinuation of study intervention and initiation of prohibited weight management treatments, or glycemic rescue therapy or prohibited glycemic therapy, which is handled by the hypothetical strategy. The potential outcome of interest is the response in the efficacy measurement if participants would remain on their randomly assigned study intervention for 72 weeks and would not initiate prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy. Dose modification and interruption will not be considered as ICEs because they are a part of the treatment condition.
- ***Population-level summary and treatment effect of interest:*** Difference in mean changes in HbA1c from baseline to Week 72 between orforglipron and placebo.

Rationale for the estimand: This estimand aims to evaluate the efficacy of orforglipron under the ideal condition that all participants would adhere to the randomly assigned study intervention without being confounded by the initiation of prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy.

1.2. Study Design

ATTAIN-J (Study GZPD) is a Phase 3, multicenter, randomized, parallel-arm, double-blind, placebo-controlled, 72-week study. The study will investigate the efficacy and safety of treatment with daily oral doses of orforglipron (6 mg, 12 mg, or 36 mg) compared with placebo when used in conjunction with a reduced calorie diet and increased physical activity for weight management in Japanese participants with obesity disease with

- body mass index (BMI) ≥27 kg/m^2 and <35 kg/m^2, and at least 2 obesity-related health problems (treated or untreated), or
- a BMI ≥35 kg/m^2 and at least 1 obesity-related health problem (treated or untreated)

according to the guidelines of the Japan Society for the Study of Obesity (JASSO) (JASSO 2022). At least 1 obesity-related health problem should be hypertension, dyslipidemia, or type 2 diabetes (T2D) (approximately 25% of participants).

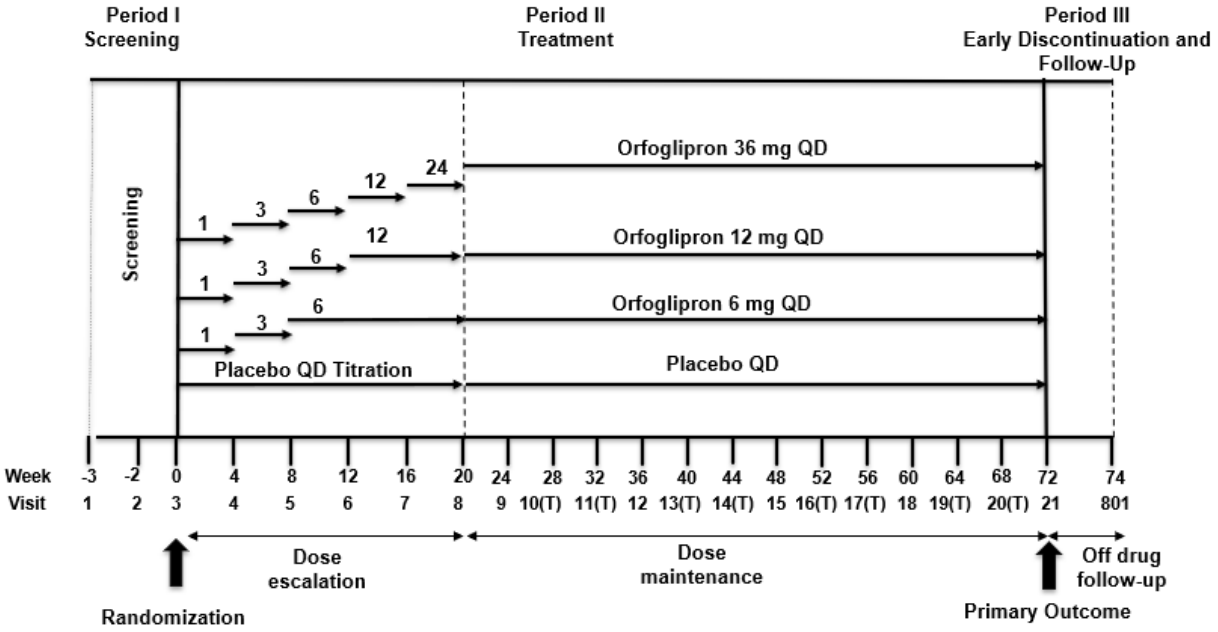
The study population will be enriched for participants who do not have T2D so that approximately 75% of the total study population will consist of participants without T2D.

Study participants will be randomized in a 1:1:1:1 ratio to once daily oral treatment with orforglipron 6 mg, 12 mg, 36 mg, or placebo. An upper limit of 70% enrollment of females will be used to ensure a sufficiently large sample of males.

The Study GZPD schema is shown in Figure GZPD.1.1.

The randomization will be stratified by:

- baseline hypertension (yes or no)
- baseline dyslipidemia (yes or no)
- baseline T2D (yes or no)
- sex (male or female), and
- BMI (≥27 to <35 kg/m^2 or ≥35 kg/m^2).



Abbreviations: QD = once daily; T= telephone visit.

Figure GZPD.1.1. Illustration of study design for Clinical Protocol J2A-JE-GZPD.

2. Statistical Hypotheses

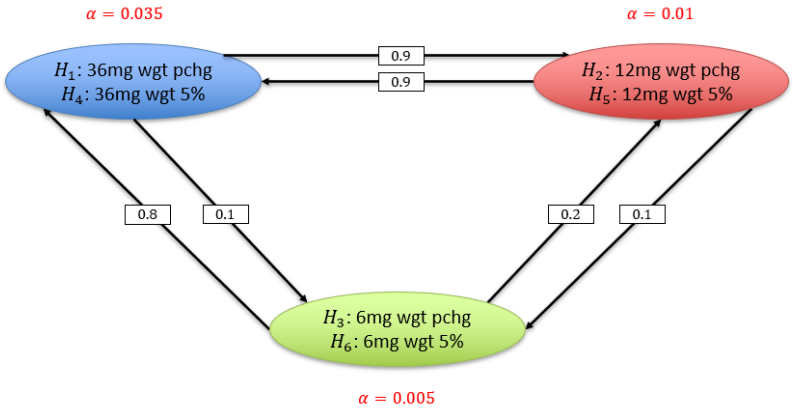
The null hypotheses corresponding to the primary objective are

- H_1: No difference in 36 mg orforglipron compared to placebo with respect to the mean percent change in body weight at Week 72.
- H_2: No difference in 12 mg orforglipron compared to placebo with respect to the mean percent change in body weight at Week 72.
- H_3: No difference in 6 mg orforglipron compared to placebo with respect to the mean percent change in body weight at Week 72.
- H_4: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction from baseline to Week 72.
- H_5: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction from baseline to Week 72.
- H_6: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction from baseline to Week 72.

2.1. Multiplicity Adjustment

Multiplicity adjusted analyses will be performed on the primary objectives in order to control the overall family-wise type I error rate at a 2-sided alpha level of 0.05. The graphical multiple testing procedure described in Bretz et al. (2009, 2011) will be used. This approach is a closed testing procedure; hence, it strongly controls the family-wise type 1 error rate across all hypotheses (Alosh et al. 2014).

Figure GZPD.2.1 illustrates the final graphical testing scheme, including testing order, interrelationships, Type I error allocation for the primary objectives, and the associated propagation. Two hypotheses corresponding to co-primary endpoints for a single dose (for example, H_1 and H_4) will be tested simultaneously under the same alpha level at every step of the sequentially rejective procedure; superiority will be declared if and only if both tests are passed. Unless otherwise specified, the treatment regimen estimand is planned to support future product registration and the efficacy estimand will be considered for use in scientific disclosures and other purposes. As these 2 types of estimands are intended for distinct purposes, no multiplicity adjustment will be made for conducting separate analyses on the same objectives. Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses outside the primary objectives.



Note: Each circle represents two hypothesis tests corresponding to co-primary endpoints, which will be tested simultaneously under the same alpha level. Superiority will be declared if and only if both tests are passed.

Figure GZPD.2.1. Graphical testing procedure.

3. Analysis Sets

For the purpose of analysis, the following populations are defined in Table GZPD.3.1, along with their attributes.

Table GZPD.3.1. Participant Analysis Sets

Participant Analysis Set	Description
Entered participants	**Definition:** all participants who sign informed consent[a]. **Purpose:** used for providing summaries for screen failures and the associated reasons. **Treatment Groups:** none. **Inferential Comparisons:** none.
Randomized participants	**Definition:** all participants who are randomly assigned a study intervention. **Purpose:** used for listings of disposition and treatment assignment summaries of disposition, demographics, historical illness, preexisting conditions, and analyses of efficacy and health outcomes. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO. **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO; OFG pooled vs. PBO.
Safety participants	**Definition:** all participants who are randomly assigned a study intervention and who take at least 1 dose of study intervention. **Purpose:** used for all safety analyses. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO. **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO.

Abbreviations: OFG = orforglipron; PBO = placebo; vs = versus.

[a] Refers to the informed consent for the study.

The following data points sets are defined for all parameters except for the glycemic-related parameters, as shown in Table GZPD.3.2.

Table GZPD.3.2. Data Points Sets for Non-Glycemic-Related Parameters

Data Points Sets	Description
Treatment regimen estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the last visit within the treatment period, regardless of study intervention discontinuation or initiation of prohibited weight management treatments. Baseline is defined in Section 4.1.1.
Efficacy estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the earliest date of study intervention discontinuation or initiation of prohibited weight management treatments. Baseline is defined in Section 4.1.1 and the prohibited weight management treatments are specified in Appendix 3 (Section 6.3) and Table GZPD.6.5.

The following data points sets are defined for all glycemic-related parameters, as shown in Table GZPD.3.3. The parameters in scope include fasting glucose and hemoglobin A1c (HbA1c).

Table GZPD.3.3. Data Points Sets for Glycemic-Related Parameters

Data Points Sets	Description
Treatment regimen estimand data points set	All data points obtained during the treatment period defined as at or after baseline, and up to the last visit within the treatment period, regardless of study intervention discontinuation or initiation of prohibited weight management treatments or glycemic rescue therapy. Baseline is defined in Section 4.1.1.
Efficacy estimand data points set	All data points obtained during the treatment period defined as at or after baseline, and up to the earliest date of discontinuation of study intervention or initiation of prohibited weight management treatments or glycemic rescue therapy. Baseline is defined in Section 4.1.1 and the prohibited weight management treatments and prohibited glycemic rescue therapy are specified in Section 6.3 and Table GZPD.6.5.

The following data points set is defined for the safety analyses.

Table GZPD.3.4. Data Points Sets for Safety Analyses

Data Points Sets	Description
Safety data points set	All data points obtained during the treatment period defined as at or after baseline, and up to the date of study withdrawal or study completion, including the follow-up period and regardless of study intervention discontinuation or initiation of prohibited weight management treatments or glycemic rescue therapy. Baseline is defined in Section 4.1.1.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in this SAP or the clinical study report (CSR). Additional exploratory data analyses may be conducted as deemed appropriate.

Details about the analyses regarding demographic and baseline characteristics, historical illnesses and preexisting conditions, treatment compliance, concomitant medications, and important protocol deviations can be found in Appendices 1 through 5 (Section 6.1, Section 6.2, Section 6.3, Section 6.4, and Section 6.5), respectively.

Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (for example, few events to justify conducting an analysis). Not all analysis described in this SAP will necessarily be included in the CSRs. Any analyses described in this SAP and not provided in the CSR would be available upon request. Not all displays will necessarily be created as a static display. Some may be incorporated into interactive display tools instead of or in addition to a static display.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and the confidence interval (CI) will be calculated at 95%, 2-sided.

The change from baseline will be calculated as the value of interest at the visit minus the baseline value. In general, percent change from baseline will be calculated as the value of change from baseline divided by the baseline value in 100% scale. For some specific predefined parameters, percent change from baseline might be calculated using a log transformation. If the baseline value is missing for a particular variable, then the change from baseline and percent change from baseline will not be calculated.

For stratification factors at baseline, a strata variable is defined for statistical modeling to consist of 4 joint levels. The strata variable includes sex (female, male) and baseline T2D (yes, no). Baseline BMI (<35 kg/m^2 or ≥ 35 kg/m^2) is also a stratification factor and will be included in statistical modeling, where appropriate, as a separate factor. Baseline T2D (yes, no), baseline hypertension (yes, no), and baseline dyslipidemia (yes, no) will be derived from using the record on the case report form (CRF) Medical History and Adverse Events Page, vital sign values, and laboratory test values. For analyses for participants with T2D or without T2D (with non-T2D) at baseline, a strata variable is replaced with sex.

Data may exist at visits where a variable was not scheduled to be collected, due to, for example, early discontinuation visits. In these situations, data from the early discontinuation visit that does not correspond to the planned collection schedule will be excluded from the mixed model for repeated measures (MMRM), analysis of covariance (ANCOVA), or logistic regression analysis, unless otherwise specified (Andersen and Millen 2013).

Handling of missing data is addressed in Section 4.1.2 and Table GZPD.4.5. Section 3 provides definitions for the participant analysis set and data points set, which includes definitions for censored data where appropriate. Handling spurious data is addressed in Section 6.5 for important protocol deviations. Protocol GZPD Section 10.1.7 addresses data quality assurance.

4.1.1. Definition of Baseline

Unless otherwise specified, the baseline for efficacy assessments is defined as the last available nonmissing measurement prior to the first dose of study intervention. In most cases, this will be the measurement recorded at Week 0 (Visit 3). If there are no doses of study intervention administered, baseline will be defined as the last available nonmissing measurement on or prior to randomization.

For safety assessments, the definition of baseline and postbaseline are specified in Table GZPD.4.1.

In cases where the measurement is taken on the same day (where the time is not collected or not reliable) as the first dose, this measurement will be used as the baseline value for data analysis. For patient-reported outcome measures, data obtained at Visit 3, regardless of the timing relative to first dose, will serve as the baseline.

The Prespecified Medical History and Adverse Events in the CRF are used for the identification of participants with obesity-related health problems (hypertension, dyslipidemia and T2D) at baseline.

Table GZPD.4.1. Baseline and Postbaseline Definitions for Safety Analyses

Analysis Type	Baseline Period	Postbaseline Period
TEAEs	Starts from informed consent date and ends prior to the first dose (typically at Week 0).	Starts after the first dose of study intervention[a] and ends at the end of the follow-up period or the date of study withdrawal.
TE abnormal laboratory values, vital signs, and ECGs	Starts from informed consent date and ends prior to the first dose (typically at Week 0). All scheduled and unscheduled measurements will be included.	Starts after the first dose and ends at the end of the follow-up period or the date of study withdrawal. All scheduled and unscheduled measurements will be included.
Change from last baseline to each postbaseline week and to last postbaseline for laboratory values, vital signs, and ECGs	When "last baseline" is used, starts from informed consent date and ends prior to the first dose (typically at Week 0).	Starts after the first dose and ends at the end of the follow-up period or the date of study withdrawal. Only scheduled visits will be included. Early termination visits that fall on the scheduled visits will be considered scheduled visits.

Abbreviations: ECG = electrocardiogram; TE = treatment-emergent; TEAE = treatment-emergent adverse events.

[a] For events occurring on the day of first dose, information collected from the case report form will be used to determine whether the event was pre- versus posttreatment if available. If the relevant information is not available, then the events will be counted as postbaseline.

4.1.2. Analysis Methods

The analysis methods are consistent with the desired estimands and the FDA guidance on "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products: Guidance Document" (FDA 2023).

4.1.2.1. Analysis Methods for Treatment Regimen Estimand

4.1.2.1.1. Method for Continuous Variables

An ANCOVA model will be used to analyze continuous measurements at Week 72 and will be guided by the treatment regimen estimand. The model will include:

- treatment group as a factor variable
- strata as a factor variable
- baseline value (of the dependent variable)
- Baseline BMI (<35 kg/m^2 or ≥ 35 kg/m^2) as a factor variable
- interaction between strata variable and treatment group, and
- interaction between baseline value and treatment group

The estimated treatment group effect and comparison between 6 mg orforglipron, 12 mg orforglipron, and 36 mg orforglipron versus placebo will be reported together with variability estimated using the robust inference (Ye et al. 2022; FDA 2023). The associated 2-sided 95% CI and corresponding p-values will also be reported. If the model fails to converge, all the interaction terms will be removed before the model fitting. The addition of interaction terms is not intended to estimate the heterogeneity effect but to provide robustness and efficiency for the

estimate of treatment comparisons on the unconditional effect. The final inference will be derived using Rubin's Rule (Rubin 1987) by combining estimates from multiple imputed datasets. The imputation procedure for creating a single imputed dataset is detailed in Table GZPD.4.2.

For some variables, both the baseline and the postbaseline values will be log-transformed before fitting the ANCOVA. The treatment group estimates will be the percent change from baseline while the treatment contrasts will be the relative change in orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg compared to the placebo. In these cases, least squares (LS) means and 95% CIs for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and relative treatment difference to placebo in percent change.

A linear contrast, averaging estimates from the individual doses, will be used to estimate the treatment effect of the pooled doses compared with placebo.

4.1.2.1.2. Primary Multiple Imputation Strategy

Discontinuation is expected to be uncommon. Participants who discontinue the study intervention (that is, discontinue study treatment) will be encouraged to continue in the study for the treatment period and follow-up period (note, the case report forms [CRFs] use "phase", which is interchangeable with the "period" in the SAP). In this section, study discontinuation refers to treatment phase discontinuation captured in the CRF.

If there are occurrences of missing data despite the best precautions, missing data should be imputed in a fashion consistent with what the values would likely have been had they been collected. In general, 5 scenarios are considered regarding the treatment journey of a trial participant relative to the primary time point (Week 72 visit), as shown in Figure GZPD.4.1.

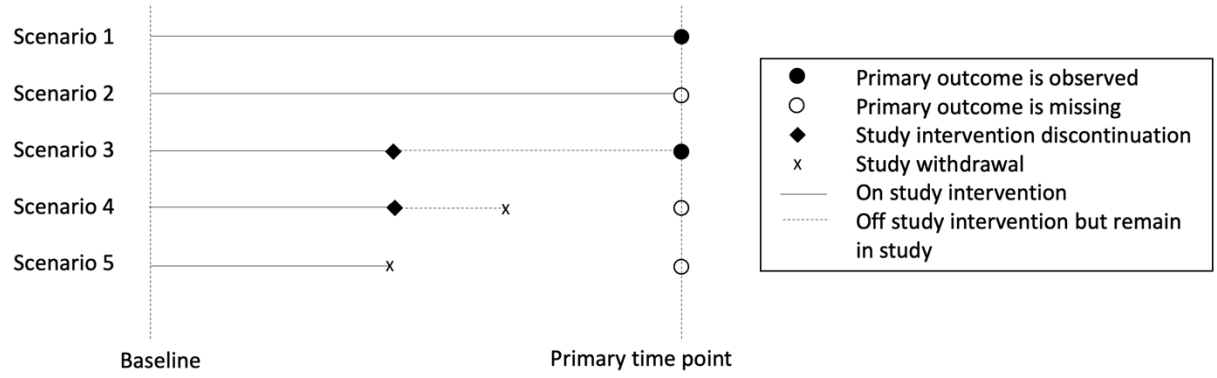


Figure GZPD.4.1. Treatment journey of a trial participant.

Scenario 5 covers the situations where a participant discontinues from the study intervention and then comes back for a study discontinuation visit for the same discontinuation reason or prior to a scheduled visit. Additional details on Scenario 5 are provided in Table GZPD.4.3.

In principle, missing data due to the permanent discontinuation of study intervention (Scenario 4 and Scenario 5A) will be imputed by treatment group using retrieved dropouts (multiple imputation retrieved dropouts [MI-RD]), namely using multiple imputation based on data retrieved from participants who permanently discontinued the study intervention but continued in the study with nonmissing measurements from the same treatment group. For scenarios 4 and 5A, if there are not enough retrieved dropouts to provide a reliable imputation model (that is, the model implemented does not converge), an alternative multiple imputation method (placebo-washout method) (Wang and Du 2023) will be used. Any missing values at baseline will be imputed in the same model when imputing the endpoint.

Table GZPD.4.2. Imputation Procedure

Scenario (Study Intervention/Treatment Period Discontinuation; Missingness Status at Endpoint)	**Methods to Handle Missing Values at Endpoint**
1. No study intervention discontinuation; no missing value	N/A
2. No study intervention discontinuation; with missing value	Use observed data, including baseline and all post-baseline visits, across all time points from Scenario 1 and 2 to impute under the MAR assumption by treatment group. There should be very few such cases in a clinical trial. No additional covariates will be included in the imputation model.
3. Study intervention discontinuation; no missing value	N/A
4. Study intervention discontinuation with treatment period discontinuation at a subsequent visit or completion of treatment period; with missing value	Missing values at the endpoint visit will be imputed through MCMC (predictive mean matching method), using baseline and the endpoint visit data from Scenario 3 (MI-RD) by treatment group. No additional covariates will be included in the imputation model.
5. Study discontinuation resulting in study intervention discontinuation; with missing value	5A: If the study discontinuation is possibly related to study intervention (see treatment period discontinuation reasons classified as Category 5A in Table GZPD.4.3), missing values at the endpoint visit will be imputed in the same way as Scenario 4.5B: If the study discontinuation is clearly due to administrative reasons not related to study intervention (see treatment period discontinuation reasons classified as Category 5B in Table GZPD.4.3), missing values at the endpoint visit will be imputed in the same way as Scenario 2.

Abbreviations: MAR = missing at random; MCMC = Markov chain Monte Carlo; MI-RD = multiple imputation using retrieved dropouts; N/A = not applicable.

Table GZPD.4.3. Categorization of Treatment Period Discontinuation Reasons

Disposition Reason	Associated Subcategories	Category
Adverse Event		5A
Assigned Treatment by Mistake		5A
Death		5A
Protocol Deviation		5A
Pregnancy		5A
Noncompliance With Study Drug		5A
Lack of Efficacy	Investigator	5A
	Study subject	5A
Withdrawal by Subject	Concern about study procedures/perceived risks	5A
	Health insurance changes	5A
	Scheduling conflicts	5A
	Subject is moving or has moved	5A
	Personal issue unrelated to trial	5A
	Due to epidemic/pandemic	5B
	Other	5A
Physician Decision	Due to epidemic/pandemic	5B
	Other	5A
Study Terminated by Sponsor		5B
Site Terminated by Sponsor		5B
Study Terminated by IRB/ERB		5B
Lost to Follow-up		5A
Other		5A

Abbreviations: ERB = ethical review board; IRB = institutional review board.

Note: For participants that discontinued the study due to emergency unblinding, the study disposition will be classified based upon the reason leading to emergency unblinding (for example, Adverse Event [5A]).

4.1.2.1.3. Multiple Imputation-Based Tipping-Point Analysis

A multiple imputation-based tipping-point (MI-TP) analysis is planned as a sensitivity analysis to explore how different patterns of percent body weight change post treatment period discontinuation by different treatment groups could impact treatment comparisons.

To start with, missing data are imputed according to the primary multiple imputation (PMI), as shown in Section 4.1.2.1.2. A penalty is then added to those imputed values at the Week 72 visit. The analysis varies the magnitude of the penalties added for both treatment groups under comparison and evaluates the impact these would have on the study conclusion. A 2-dimensional space of penalties will be assessed for orforglipron 36 mg, orforglipron 12 mg, or orforglipron 6 mg and placebo. MI-TP aims to evaluate the robustness of the superiority claim to the assumptions of using the observed data to impute the missing body weight in all treatment groups.

4.1.2.1.4. Method for Binary Variables

The binary outcomes will be analyzed with following the procedure:

- Using the same imputation strategy – PMI as in Section 4.1.2.1.2 on the underlying continuous endpoint that determines the binary value. For example, the continuous body weight values at Week 72 visit will be imputed first.
- Transform the observed and imputed continuous value to the binary value, for example, convert the continuous body weight values at Week 72 visit to whether the corresponding percent change from baseline meets a certain threshold (Ma et al. 2022).
- Fit a logistic regression model to the data with the following terms:
    - treatment group as a factor variable (for pooled orforglipron analyses, treatment group will be a 2-level factor for pooled orforglipron and placebo)
    - strata as a factor variable
    - baseline BMI (<35 kg/m^2 or ≥ 35 kg/m^2) as a factor variable, and
    - continuous baseline value (of the dependent variable)
    - interaction between treatment group and strata, and
    - interaction between treatment group and baseline value.

- Provide estimates and inferences of unconditional treatment effects defined by the risk difference and/or relative risk based on the delta-method using the formula provided (Ye et al. 2023).
- Derive the final inference using Rubin's rule which combines estimates from multiple imputed datasets. The imputation procedure for creating a single imputed dataset is detailed in Table GZPD.4.2.

The statistical testing for risk difference will be used as the primary approach for treatment comparisons.

4.1.2.1.5. Analysis Methods for Time-to-Event Variables

For time-to-event analysis, the log-rank test stratified on the strata variable will be conducted for the comparison between treatment groups. A Cox proportional hazards model conditioned on the strata variable and with treatment group as a factor, will be used as the primary analysis to estimate the hazard ratio (HR) between the treatment groups and the corresponding CI.

For time-to-event analyses, participants who have not experienced the event of interest will be censored at the participant's end of follow-up. The censoring date will be the earliest of

1. date of death
2. date of study completion/withdrawal, and
3. the last confirmed visit date before participant was lost-to-follow-up.

The Kaplan-Meier method will be used to estimate the cumulative event curve over time. Counts and proportions of participants who experience an event will be calculated by treatment group.

For time-to-event analysis, if a participant experiences the event of interest, then time-to-event for that specific event of interest will be the number of days between the date of randomization and the onset date of the event plus 1 day. If a participant does not experience the event, then time-to-event for that specific event of interest will be the number of days between the date of randomization and the date of the participant's end of follow-up plus 1 day. If a participant

experiences multiple events the date of the first event will be used, unless otherwise specified. In rare cases where randomization occurred prior to the Visit 3 date, the Visit 3 date will be used instead.

4.1.2.2. Analysis Methods for Efficacy Estimand

4.1.2.2.1. Method for Continuous Variables

A maximum likelihood-based MMRM will be used to analyze continuous measurements at each postbaseline visit, guided by the efficacy estimand. Missing data should be minimized at the best precaution. The hypothetical strategy will be used to handle ICEs. For MMRM analysis, only data collected before the occurrence of any ICEs will be used in the MMRM analysis. Through the MMRM, the potential efficacy measures (after the ICEs) will be implicitly imputed as if participants did not have ICEs. Participant dose modification and interruption will not disqualify their measurements from being included into the model.

The MMRM model will include the following terms:

- treatment group as a factor variable
- visit as a factor variable
- strata as a factor variable
- baseline BMI (<35 kg/m^2 or ≥ 35 kg/m^2) as a factor variable
- baseline value (of the dependent variable)
- interaction between treatment group, strata variable, and visit, and
- interaction between treatment group, baseline value, and visit.

The estimated treatment group effect and comparison between 6 mg orforglipron, 12 mg orforglipron, or 36 mg orforglipron and placebo at the scheduled visits will be reported together with the variability estimated using robust inference (Wang and Du 2023). The sandwich estimator (Diggle et al. 1994) for the variance-covariance matrix will be used. The addition of 3-way interaction terms is not intended to estimate the heterogeneity effect but to provide robustness and efficiency for the estimate of treatment comparisons on the unconditional effect. The associated 2-sided 95% CI and corresponding p-values will also be reported. An unstructured covariance matrix by each treatment group will be used to model the within-participant errors, assuming heteroscedasticity and the measurements for different participants are independent. If the model fails to converge, the model will be simplified to include the following terms, removing 3-way interactions and the heteroscedasticity assumption:

- treatment group as a factor variable (for pooled orforglipron analyses, treatment group will be a 2-level factor for pooled orforglipron and placebo)
- visit as a factor variable
- strata as a factor variable
- baseline BMI (<35 kg/m^2 or ≥ 35 kg/m^2) as a factor variable
- baseline value (of the dependent variable)interaction between treatment group and visit
- interaction between strata variable and visit, and
- interaction between baseline value and visit.

If the model still fails to converge, the following covariance structures will be tested in order for the simplified model:

- heterogeneous Toeplitz
- heterogeneous autoregressive
- heterogeneous compound symmetry
- homogeneous Toeplitz
- homogeneous autoregressive, and
- homogeneous compound symmetry.

The first covariance structure that converges will be used.

For some variables, both the postbaseline response variables and baseline variable will be log-transformed before fitting the MMRM. The treatment group estimates will be the percent change from baseline while the treatment contrasts will be the relative change in orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg compared to the placebo (%) over the scheduled visits.

A linear contrast, averaging estimates from the individual doses, will be used to estimate the treatment effect of the pooled doses compared with placebo.

4.1.2.2.2. Method for Binary Variables

The binary outcomes will be analyzed with the following procedure:

- Impute the missing continuous-valued measurements at the scheduled visits using the randomized participants set with the efficacy estimand data points set assuming missing at random with multiple imputation.
- At the visit of interest, transform the observed and imputed continuous value to the binary value, for example, convert the continuous body weight value at a visit to whether the corresponding percent change from baseline meets a certain threshold.
- Fit a logistic regression model to the transformed binary data with the following terms:
    - treatment group as a factor variable (for pooled orforglipron analyses, treatment group will be a 2-level factor for pooled orforglipron and placebo)
    - strata as a factor variable
    - baseline BMI (<35 kg/ m^2 or ≥ 35 kg/ m^2) as a factor variable
    - continuous baseline value (of the dependent variable, if any underlying continuous variable)
    - interaction between treatment group, strata variable, and visit, and
    - interaction between treatment group, baseline value, and visit.

- For the visit of interest, provide the estimates and inferences of unconditional treatment effects defined by the risk difference and/or risk ratio based on the delta-method using the formula provided (Ye et al. 2023).
- Derive the final inference using Rubin's Rule by combining estimates from multiple imputed datasets.

The statistical testing for risk difference will be considered the primary method for treatment comparisons.

4.1.2.2.3. Analysis Methods for Time-to-Event Variables

Similar methodology will be used as described in Section 4.1.2.1.5, with the exception of the censoring rules.

For time-to-event analyses, participants who have not experienced the event of interest will be censored at the participant's end of follow-up. The censoring date will be the earliest of

1. date of death
2. date of initiation of prohibited weight management treatments or glycemic therapy (for glycemic related parameters, date of initiation of glycemic therapy is considered)
3. date of discontinuation of study intervention, or
4. the last confirmed visit date before participant was lost-to-follow-up.

4.1.2.3. Analysis Methods for Safety

In general, safety assessments will be guided by an estimand comparing safety of orforglipron doses with placebo irrespective of adherence to study intervention. Thus, safety analyses will be conducted using the safety participants based on data observed during treatment period plus the posttreatment follow-up period, that is, the period from first dose of treatment to the end of the follow-up visit or the date of study withdrawal. It is noted that for some safety endpoints, such as, treatment discontinuation due to adverse event (AE) and so on, the analyses will be based on data observed during treatment period only.

Fisher's exact test will be used for treatment comparisons of percentages. The risk differences and 95% CIs will be provided.

For selected continuous safety measures, unless otherwise specified, treatment group differences of mean change or percent mean change from baseline at all scheduled visits will be assessed via an MMRM using maximum likelihood, which will include the following terms:

- treatment group as a factor variable
- visit as a factor variable
- baseline value (of the dependent variable), and
- interaction between visit and treatment group.

If the data does not warrant the MMRM model, then an ANCOVA model with treatment group as a fixed effect and the continuous baseline value as a covariate will be used.

No explicit imputation will be conducted for safety measures. Some parameters (such as urinary albumin to creatinine ratio, p-amylase, and lipase) may be log-transformed before fitting the MMRM as specified in Section 4.6

The Kaplan-Meier product limit method will be used to estimate the cumulative event-free survival rates over time for the time-to-event analyses. An unstratified Cox proportional hazards regression analysis will be used to compare hazard rates among treatments. An unstratified

log-rank test will be used to calculate p-values. Time-to-event for the specific safety event of interest will be calculated from the date of first dose.

The logistic regression model will be used for hypoglycemia incidence for postbaseline treatment comparisons. The model will include the fixed effects of treatment and strata variable. Where necessary, the rate of events will be analyzed using a generalized linear mixed-effects model assuming the number of events follow a negative binomial distribution and treatment group as a fixed effect. The logarithm of days during the analysis interval will be adjusted as an offset to account for the possible unequal treatment duration of follow-up between participants.

Some safety analyses may be conducted after excluding data after the initiation of new antihyperglycemic medications for participants with baseline T2D.

4.2. Participant Dispositions

The participant dispositions for the screening period, the study intervention, the treatment period and/or the follow-up period will be collected in CRFs with the corresponding primary reason. The study completion for a participant is defined as the participant completing both the treatment period and the follow-up period, regardless of completion of study intervention.

The planned listings and summary tables for dispositions are provided in Table GZPD.4.4. No inferential analysis will be performed.

Time-to-study intervention discontinuation will be calculated from the date of first dose. Time-to-study discontinuation will be calculated from the date of randomization.

Table GZPD.4.4. Listings and Summary Tables Related to Dispositions

Analysis	Population/Period
Summary of disposition (prior to randomization)	Entered participants
Patient allocation by center/site	Entered participants
Summary of study and study intervention disposition	Randomized participants/TP + FP (TP for study intervention disposition)
Kaplan-Meier plot of time to study discontinuation	Randomized participants/TP + FP
Kaplan-Meier plot of time to study intervention discontinuation	Safety participants/TP
Kaplan-Meier plot of time to study intervention discontinuation due to AEs	Safety participants/TP
Listing of randomization	Randomized participants
Listing of randomized participants not administering study intervention	Randomized participants
Listing of randomized participants who were discontinued from the study intervention due to inadvertent enrollment	Randomized participants
Listing of study and study intervention disposition	Randomized participants

Abbreviations: AE = adverse event; FP = follow-up period; TP = treatment period.

4.3. Primary Endpoints/Estimands Analysis

4.3.1. Definition of Endpoint(s)

The primary efficacy endpoints are percent change in body weight and achieving 5% or more body weight reduction from baseline to Week 72.

As specified in Table GZPD.4.5, the percent change in body weight is defined as:

(postbaseline body weight [kg] – baseline body weight [kg]) / baseline body weight [kg] × 100%

and achieving 5% or more body weight reduction is defined as:

Response = yes if at least a 5% reduction in body weight from baseline, that is, percent change from baseline in body weight ≤ -5%.

4.3.2. Main Analytical Approach

The analytical approaches for percent change in body weight at 72 weeks are specified in Section 4.1.2.1.1 and Section 4.1.2.1.2 (ANCOVA with PMI) and Section 4.1.2.2.1 (MMRM).

The analytical approaches for achieving 5% or more body weight reduction from baseline to Week 72 are specified in Section 4.1.2.1.4 (logistic regression with PMI) and Section 4.1.2.2.2 (logistic regression with multiple imputation [MI]).

4.3.3. Sensitivity Analyses

The sensitivity analyses for percent change in body weight at 72 weeks are specified in Section 4.1.2.1.3 (ANCOVA with MI-TP).

4.3.4. Supplementary Analyses

No supplementary analyses are planned for the primary endpoint.

4.4. Secondary Endpoints/Estimand Analysis

Table GZPD.4.5. includes the description and derivation of all the primary and secondary endpoints for the efficacy measures and patient-reported outcomes.

Table GZPD.4.6. provides the detailed analyses including endpoint, estimand, data points set, method and imputation, population, and time point for the primary and secondary endpoints.

PK/PD analyses will be documented in a separate document.

Table GZPD.4.5. Description and Derivation of Primary and Secondary Endpoints for Efficacy Measures and Patient Reported Outcomes

Measure	Description	Variable	Derivation/Comment	Handling Missing Components[a]
Body weight	Per Protocol GZPD Section 10.7: • Body weight measurements should be done in a consistent manner using a calibrated electronic scale capable of measuring weight in kilograms to 1 decimal place. • All weights for a given participant should be measured using the same scale, whenever possible, at approximately the same time in the morning after evacuation of bladder contents. • Body weight will be measured in the fasting state at all visits. If the participant is not fasting, the participant should be called in for a new visit within the visit window to have the fasting body weight measured.	Body weight (kg)	As measured	M1
		CHG in body weight (kg)	Calculated as: postbaseline body weight (kg) – baseline body weight (kg)	M2
		%CHG in body weight (%)	Calculated as: (postbaseline body weight [kg] – baseline body weight [kg]) / baseline body weight [kg] × 100 (%)	M2
		≥x% body weight reduction from baseline where, x = 5, 7, 10, 15, 20	Response = yes if at least a x% reduction in body weight from baseline, that is, percent change from baseline in body weight ≤ -x%	M2
BMI	BMI: Round to 1 decimal point. For example, a BMI of 26.6 kg/m^2 should **not** be rounded to 27.0 kg/m^2.	BMI (kg/m^2)	BMI will be calculated as: Weight(kg)/(Height[m])^2	Missing if weight or height is missing.
		CHG in BMI (kg/m^2)	Calculated as: postbaseline BMI (kg/m^2) – baseline BMI (kg/m^2)	M2
		BMI target value of <27 kg/m^2	Response = yes if the postbaseline BMI value <27 kg/m^2	Missing if postbaseline value is missing.
		BMI target value of <25 kg/m^2	Response = yes if the postbaseline BMI value <25 kg/m^2	Missing if postbaseline value is missing.
Waist circumference	Per Protocol GZPD Section 10.7: • Waist circumference should be measured in the horizontal plane and at the umbilicus level according to JASSO guidelines.	Waist circumference (cm)	The average of 2 measures.	If there is at least one measurement, take the average of all nonmissing

	• Measurements should be taken at the end of a normal expiration using a non-stretchable measuring tape. The tape should lie flat against the skin without compressing the soft tissue. • The waist circumference should be measured twice, rounded to the nearest 0.5 cm. The measuring tape should be removed between the 2 measurements. Both measurements will be recorded in the CRF. If the difference between the 2 measurements exceeds 1 cm, this set of measurements should be discarded and the 2 measurements repeated.			measurements; otherwise, set it to be missing.
		CHG in waist circumference (cm)	Calculated as: postbaseline Waist circumference (cm) – baseline Waist circumference (cm)	M2
Blood pressure	Per Protocol GZPD Section 10.7: • Have the participant sit quietly for about 5 minutes before vital signs measurements are taken. • Take 3 measurements from the same arm, preferably the nondominant arm. • Measure the recordings at least 1 minute apart. • Blood pressure must be taken with an automated blood pressure instrument.	SBP (mm Hg)	The average of 3 measures.	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing.
		CHG in SBP (mm Hg)	Calculated as: postbaseline SBP (mm Hg) – baseline SBP (mm Hg)	M2
		DBP (mm Hg)	The average of 3 measures.	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing.
		CHG in DBP (mm Hg)	Calculated as: postbaseline DBP (mmHg) – baseline DBP (mmHg)	M2
Lipid parameter	HDL-cholesterol, triglycerides, and total cholesterol will be assayed by a Lilly designated laboratory. LDL-cholesterol, non-HDL-cholesterol, and VLDL-cholesterol will be calculated by a Lilly	Triglycerides	As provided. Log transformation before the analysis.	M1

	designated laboratory.	%CHG in triglycerides (%)	Calculated as: log(postbaseline triglyceride) – log(baseline triglyceride) then will transform back to percent change.	M2
		Total cholesterol	As provided. Log transformation before the analysis.	M1
		%CHG in total cholesterol (%)	Calculated as: log(postbaseline total cholesterol) – log(baseline total cholesterol) then will transform back to percent change.	M2
		Non-HDL-cholesterol	As provided. Log transformation before the analysis.	M1
		%CHG in non-HDL-cholesterol (%)	Calculated as: log(postbaseline non-HDL-cholesterol) – log(baseline non-HDL-cholesterol) then will transform back to percent change.	M2
		LDL-cholesterol	As provided. If triglycerides are >400 mg/dL, the direct LDL will be directly measured. Log transformation before the analysis.	M1
		%CHG in LDL-cholesterol (%)	Calculated as: log(postbaseline LDL-cholesterol) – log(baseline LDL-cholesterol) then will transform back to percent change.	M2
		HDL-cholesterol	As provided. Log transformation before the analysis.	M1
		%CHG in HDL-cholesterol (%)	Calculated as: log(postbaseline HDL-cholesterol) – log(baseline HDL-cholesterol) then will transform back to percent change.	M2
		VLDL-cholesterol	As provided. Log transformation before the analysis.	M1
		%CHG in VLDL-cholesterol (%)	Calculated as: log(postbaseline VLDL-cholesterol – log(baseline VLDL-cholesterol) then will transform back to percent change.	M2

Glycemic control	Fasting glucose will be performed by the central laboratory and will be assayed by a Lilly-designated laboratory.	Fasting glucose (mg/dL and mmol/L)	As provided.	M1
		CHG fasting glucose (mg/dL and mmol/L)	Calculated as: postbaseline fasting glucose – baseline fasting glucose	M2
	HbA1c will be performed by the central laboratory and will be assayed by a Lilly-designated laboratory.	HbA1c (% and mmol/mol)	As provided.	M1
		CHG HbA1c (% and mmol/mol)	Calculated as: postbaseline HbA1c – baseline HbA1c	M2
	Fasting C-peptide will be performed by the central laboratory and will be assayed by a Lilly-designated laboratory.	Fasting C-peptide (nmol/L, µg/L)	As provided	M1
		%CHG in fasting C-peptide (%)	Calculated as: log(postbaseline fasting C-peptide) – log(baseline fasting C-peptide) then will transform back to percent change.	M2
	Participants must collect two 7-point SMBGs on 2 nonconsecutive days before assigned visits. A 7-point SMBG consists of measurements before and 2 hours after each of 3 main meals within the same day and at bedtime, that is: • Point 1: before morning meal fasting • Point 2: 2 hours after morning meal • Point 3: before midday meal • Point 4: 2 hours after midday meal • Point 5: before evening meal • Point 6: 2 hours after evening meal • Point 7: bedtime The participant needs to collect these measurements within 14 days before the assigned visits. If more than two 7-point SMBG profiles are available, the 2 most recent nonconsecutive profiles should be used.	Point x SMBG (mg/dL) x = 1, 2, …,7	The average of 2 measures in the 2 nonconsecutive days at each time point.	If there is at least one measurement, take the average of all nonmissing measurements; otherwise, set it to be missing.
		CHG in point x SMBG (mg/dL)	Calculated as: postbaseline point × SMBG (mg/dL) – baseline point × SMBG (mg/dL)	M2
		Mean of all 7-point SMBG (mg/dL)	Calculated as: (average of point 1 + average of point 2 +… + average of point 7)/7	If there are at least 4 measurements, take the average of all nonmissing measurements; otherwise, set it to be missing.

		CHG in mean of all 7-point SMBG (mg/dL)	Calculated as: postbaseline mean of all 7-point SMBG (mg/dL) – baseline mean of all 7-point SMBG (mg/dL)	M2
		Mean of pre-meal SMBG (mg/dL)	Calculated as: (average of point 1 + average of point 3 + average of point 5)/3	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing.
		CHG in mean of pre-meal SMBG (mg/dL)	Calculated as: postbaseline mean of pre-meal SMBG (mg/dL) – baseline mean of pre-meal SMBG (mg/dL)	M2
		Mean of 2-hour post-meal SMBG (mg/dL)	Calculated as: (average of point 2 + average of point 4 + average of point 6)/3	If there is at least 1 measurement, take the average of all non-missing measurements; otherwise, set it to be missing.
		CHG in post-meal SMBG (mg/dL)	Calculated as: postbaseline mean of post-meal SMBG (mg/dL) – baseline mean of post-meal SMBG (mg/dL)	M2
	Delayed progression to T2D is evaluated by time to onset of T2D (only nonT2D at baseline).	Time to onset of T2D	Calculated as: (date of onset of T2D events – date of first dose of study intervention) + 1 day Onset of T2D will be based on CRF.	Censored if no postbaseline event of T2D was observed.
	Change in glycemic category (normo-glycaemia, borderline type, T2D)	Glycemic category	Normo-glycemia: FG <110 mg/dL (6.1 mmol/L) and HbA1c <6.0% (42 mmol/mol) Borderline type: FG <126 mg/dL (7.0 mmol/L) and HbA1c <6.5% (48	M2

			mmol/mol), and NOT normo-glycemia T2D: FG ≥126 mg/dL (7.0 mmol/L) or HbA1c ≥6.5% (48 mmol/mol)	
Insulin sensitivity	Per Protocol GZPD Section 10.2: Fasting insulin will be assayed by Lilly-designated laboratory.	%CHG in fasting insulin (%)	Calculated as: log(postbaseline fasting insulin) – log(baseline fasting insulin) then will transform back to percent change.	M2
Renal function	Per Protocol GZPD Section 10.2: UACR and eGFR will be assayed by a Lilly-designated laboratory.	UACR	As provided. Log transformation before the analysis.	M1
		%CHG in UACR (%)	Calculated as: log(postbaseline UACR) – log(baseline UACR) then will transform back to percent change.	M2
		eGFR JSN Creatinine (mL/min/1.73m^2)	As provided as: eGFR JSN Creatinine	M1
		CHG in eGFR JSN Creatinine (mL/min/1.73m^2)	Calculated as: postbaseline eGFR JSN Creatinine– baseline eGFR JSN Creatinine	M2
		eGFR JSN Cystatin C (mL/min/1.73m^2)[b]	As provided as: eGFR JSN Cystatin C Calculated as: $104 \times \text{Cystatin C (mg/dL)}^{(-1.019)} \times 0.996^{(\text{Age (years)})}$ (× 0.929 for female) - 8	M1
		CHG in eGFR JSN Cystatin C (mL/min/1.73m^2)	Calculated as: postbaseline eGFR JSN Cystatin C – baseline eGFR JSN Cystatin C	M2
Inflammatory biomarker	Per Protocol GZPD Section 10.2: The hsCRP will be assayed by a Lilly-designated laboratory.	hsCRP (mg/L)	As provided.	M1
		%CHG in hsCRP (mg/L)	Calculated as: log(postbaseline hsCRP (mg/L)) – log(baseline hsCRP (mg/L)) then will transform back to percent change	M2
Biomarker	Per Protocol GZPD Section 10.2: The uric acid, adiponectin (µg/dL) (both high molecular weight and total adiponectin) and leptin	Uric acid (mg/dL)	As provided	M1
		CHG in uric acid (mg/dL)	Calculated as:	M2

	(ng/dL) will be performed by the central laboratory and will be assayed by a Lilly-designated laboratory.		postbaseline uric acid (mg/dL) – baseline uric acid (mg/dL)	
		Adiponectin (µg/dL) (both high molecular weight and total adiponectin)	As provided.	M1
		CHG in adiponectin (µg/dL) (both high molecular weight and total adiponectin)	Calculated as: postbaseline adiponectin (µg/dL) – baseline adiponectin (µg/dL)	M2
		Leptin (ng/dL)	As provided.	M1
		CHG in leptin (ng/dL)	Calculated as: postbaseline leptin (ng/dL) – baseline leptin (ng/dL)	M2
FFAs	Per Protocol GZPD Section 10.2: FFAs will be performed by the central laboratory and will be assayed by a Lilly-designated laboratory.	FFAs (mmol/L)	As provided	M1
		%CHG in FFAs (mmol/L)	Calculated as: log(postbaseline FFAs (mmol/L)) – log(baseline FFAs (mmol/L)) then will transform back to percent change	M2
HOMA	The HOMA2 B estimates steady state beta cell function.	HOMA2 B (insulin)	Calculated from the link[a] using fasting glucose and insulin laboratory values	Missing if fasting glucose or insulin is missing.
		Percent change from baseline in HOMA2 B (insulin)	Calculated as: log [postbaseline HOMA2 B (C-peptide)] – log [baseline HOMA2 B (C-peptide)]	M2
		HOMA2 B (C-peptide)	Calculated from the link[c] using fasting glucose and C-peptide laboratory values	Missing if fasting glucose or C-peptide is missing
		Percent change from baseline in	Calculated as:	M2

		HOMA2 B (C-peptide)	log [postbaseline HOMA2 B (C-peptide)] – log [baseline HOMA2 B (C-peptide)]	
	The HOMA2 S estimates insulin resistance	HOMA2 S (insulin)	Calculated from the link[a] using fasting glucose and insulin laboratory values	Missing if fasting glucose or insulin is missing.
		Percent change from baseline in HOMA2 S (insulin)	Calculated as: log [postbaseline HOMA2 S (insulin)] – log [baseline HOMA2 S (insulin)]	M2
		HOMA2 S (C-peptide)	Calculated from the link[c] using fasting glucose and C-peptide laboratory values	Missing if fasting glucose or C-peptide is missing
		Percent change from baseline in HOMA2 S (C-peptide)	Calculated as: log [postbaseline HOMA2 S (C-peptide)] – log [baseline HOMA2 S (C-peptide)]	M2
SF-36	The Short Form 36 Version 2 Health Survey (SF-36v2) will assess health-related quality of life. The SF-36v2 acute form, 1-week recall version is a 36-item generic, participant-completed measure designed to assess the following 8 domains: • Physical functioning • Role-physical • Bodily pain • General health • Vitality • Social functioning • Role-emotional, and • Mental health. The Physical Functioning domain assesses limitations due to health now. The Physical	SF-36 domain scores and SF-36 component scores	Per copyright owner, the Quality Metric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will recalibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (T-scores) using the 1-week recall period. This entails exporting the patient data in a CSV or tab-delimited file for import, generation of the SF-36 scores and reports, and export of the calculated scores in a CSV or tab-delimited file.	Missing data handling offered by SF-36 software will be used. “Maximum Data Recovery” will be selected for Missing Score Estimator in the software.

	functioning domain does not have a recall period, although the remaining domains assess functioning "in the past week." Each domain is scored individually and information from these 8 domains is further aggregated into 2 health component summary scores, a Physical Component Summary, and a Mental Component Summary. Items are answered on Likert scales of varying lengths (3-point, 5-point, or 6-point scales). Scoring of each domain and both summary scores are norm-based and presented in the form of T-scores, with a mean of 50 and SD of 10. Higher scores indicate better levels of function and/or better health (Maruish 2011).	SF-36 change from baseline for domains and component scores (PCS and MCS)	Calculated as: postbaseline SF-36 score – baseline SF-36 score	M2
PGI-S	The PGI-S Physical Function Weight scale is designed to assess the participants' overall perception of their condition. This is a single global item that asks participants to rate how their weight limited their ability to perform physical activities in the past 7 days.	PGI-S rating	As assessed. The responses are on a 5-point scale ranging from "not at all limited" to "extremely limited."	M1
PGI-C	The PGI-C Physical Function Weight scale is designed to assess the participants' overall perception of the efficacy of treatment. This is a single global item that asks participants to rate the overall change in their ability to perform physical activities due to their weight since starting the study intervention.	PGI-C rating	As assessed. The responses are based on a 5-point scale ranging from "much better" to "much worse."	M1
IWQOL-Lite-CT	The IWQOL-Lite-CT Version (Kolotkin et al. 2017, 2019) is a 20-item, obesity-specific, patient-reported outcome instrument developed for use in weight management clinical studies. The IWQOL-Lite-CT assesses 2 primary domains of obesity-and health-related quality of life: Physical (7 items) and Psychosocial (13 items), and a 5-item subset of the Physical domain, the Physical Function composite. Items in the Physical Function composite describe physical impacts	Physical domain score and psychosocial domain score	Physical domain includes Items 1-5, 16, 17; Psychosocial domain includes Items 6-8, 9-15, 18, 19, and 20.	The minimum requirements for nonmissing responses: Psychosocial: 7 of 13 items, Physical: 4 of 7 items, otherwise considered

	related to general and specific physical activities. All items are rated on either a 5-point frequency ("never" to "always") scale or a 5-point truth ("not at all true" to "completely true") scale. The 2 domain scores, 1 composite score and total score range from 0 to 100 with higher scores indicating greater functioning, derived as below. Raw scores for each subscale are computed if a minimum of 50% of the items for that subscale are nonmissing, and for the IWQOL-Lite-CT total score if a minimum of 75% of all items are nonmissing. If the minimum number of items is answered for a composite, then • compute the average of the valid nonmissing responses corresponding to the items in the total or each domain/composite will be calculated (1 = "never" or "not at all true" and 5 = "always" or "completely true"). • the score will be then calculated by transforming the raw score to the 0 (worst) to 100 (best) metric using the following formula for every participant at each time point: $100(S_{max} - C_{avg})/(S_{max} - S_{min})$ ○ C_{avg} is the raw average score of all nonmissing item responses in the composite; this average must be a number between 1 and 5, inclusive. ○ S_{max} is the maximum possible raw score value (that is, 5). ○ S_{min} is the minimum possible raw score value (that is, 1). ○ Inserting the maximum and minimum possible score values, the formula is reduced to: $100(5 - C_{avg})/4$.			missing.
		CHG in physical domain score and psychosocial domain score	Calculated as: postbaseline score – baseline score	M2
		Physical function composite score	Include Items 1-3, 16, and 17.	The minimum requirements for nonmissing responses: 3 of 5 items, otherwise considered missing.
		CHG in physical function composite score	Calculated as: postbaseline score – baseline score	M2
		Total score	Items 1-20	The minimum requirements for nonmissing responses: 15 of 20 items, otherwise considered missing.
		CHG in total score	Calculated as: postbaseline score – baseline score	M2
EQ-5D-5L and VAS	The EQ-5D-5L (EuroQol Research Foundation 2019) is a standardized, 5-item, self-administered	EQ-5D-5L item scores	Five health profile dimensions, each dimension has 5 levels:	Each dimension is M1. Note:

	instrument for use as a measure of health outcome. It provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care as well as population health surveys. The EQ-5D-5L assesses 5 dimensions of health: • Item 1: mobility • Item 2: self-care • Item 3: usual activities • Item 4: pain/discomfort • Item 5: anxiety/depression Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The VAS records the respondent's self-rated health on a vertical VAS where the endpoints are labeled as "best imaginable health state" (100) and "worst imaginable health state" (0).		1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems, and 5 = extreme problems. It should be noted that the numerals 1-5 have no arithmetic properties and should not be used as a primary score.	Missing value can be coded as "9."
		EQ-5D-5L JP population-based utility index	EQ-5D-5L health states are converted into a single index "utility" score using a scoring algorithm based on public preferences. Uses the concatenation of the value of each EQ-5D-5L dimension score in the order: item 1; item 2; item 3; item 4; item 5. Derive EQ-5D-5L JP population-based index score according to the link[d] by using the JP algorithm to produce a patient-level index score between -0.025 and 1.0 (continuous variable).	If any of the items is missing or coded as 9, the index score is missing.
		CHG in EQ-5D-5L utility index	Calculated as: postbaseline utility index score - baseline utility index score	M2
		EQ-5D VAS	Range from 0 = "worst imaginable health state" to 100 = "best imaginable health state" Note: a higher value indicates better health state.	M1
		CHG in EQ-5D-5L VAS	Calculated as: postbaseline VAS score – baseline VAS score	M2
PROMIS	The PROMIS Short Form v1.0 Sleep Disturbance 8b assesses self-reported perceptions of sleep quality, sleep depth, and restoration associated with sleep, including perceived difficulties and concerns with getting to sleep or staying asleep, as well as perceptions of the adequacy of and satisfaction with sleep. The PROMIS Short Form	PROMIS individual and total scores	T-scores are generated using response pattern scoring with a mean of 50 and a SD of 10[e].	Each item is M1
		CHG in individual and total scores	Calculated as: postbaseline score – baseline score	M2

	v1.0 Sleep Disturbance 8b has a recall period of 7 days and each of its 8 items are rated on a 5-point scale ranging from "not at all" to "very much", "never" to "always," or "very poor" to "very good." Individual item scores are totaled to obtain a raw score, with higher scores indicating more sleep disturbance.			
Improvement of comorbidities	Improvement in hypertension is evaluated by SBP and DBP (only for participants with hypertension at baseline).	Hypertension improvement	Response = postbaseline SBP<130 mm Hg and DBP<80 mm Hg	Missing if all postbaseline value is missing
	Improvement in dyslipidemia is evaluated by LDL-C, TG, or HDL-cholesterol (only for participants with dyslipidemia at baseline).	Dyslipidemia improvement	Response = yes if (all LDL-C, TG, or HDL-cholesterol that correspond to the baseline criteria, i.e. LDL-cholesterol ≥160 mg/dL, TG ≥ 150 mg/dL or HDL-cholesterol <40 mg/dL, have improved at postbaseline, that means LDL-cholesterol < 140 mg/dL, TG < 150 mg/dL and HDL-cholesterol ≥40 mg/dL)	Missing if baseline or postbaseline value is missing for all measures.
		LDL-cholesterol improvement	Response = yes if (the postbaseline LDL-cholesterol < 140 mg/dL and baseline LDL-cholesterol ≥ 140 mg/dL)	M2
		Triglyceride improvement	Response = yes if (TG < 150 mg/dL and baseline TG ≥ 150 mg/dL)	M2
		HDL-cholesterol improvement	Response = yes if (HDL-cholesterol ≥40 mg/dL and baseline HDL-cholesterol <40 mg/dL)	M2
	Improvement in T2D is evaluated by HbA1c (only for participants with T2D at baseline).	T2D (HbA1c target value of <6.5%)	Response = yes if the postbaseline HbA1c value <6.5%	Missing if postbaseline value is missing
		HbA1c target value of <5.7%	Response = yes if the postbaseline HbA1c value <5.7%	Missing if postbaseline value is missing
		HbA1c target value of <7%	Response = yes if the postbaseline HbA1c value <7%	Missing if postbaseline value is missing

	The number of improvements of obesity-related health problems (hypertension, dyslipidemia, and T2D).	Improvements of obesity-related health problems	The number of improvements of obesity-related health problems (hypertension, dyslipidemia, and T2D)	Missing if all improvement is missing
	Level of medication intensity (decrease, no change, increase)	Change in anti-hypertensive medication	As provided.	M1
		Change in lipid-lowering medication	As provided.	M1
		Change in concomitant oral anti-glycemic medication	As provided.	M1
VAT, SAT	Single slice imaging will be performed at the umbilicus level in a supine position using CT at each site in about 40% of total participants. The assessments will include VAT, SAT, and VAT/SAT ratio and will be calculated using software. Measurements of the adipose area will be performed centrally.	VAT (cm^2)	As provided.	M1
		CHG VAT (cm^2)	Calculated as: postbaseline VAT – baseline VAT	M2
		SAT (cm^2)	As provided.	M1
		CHG SAT (cm^2)	Calculated as: postbaseline VAT – baseline VAT	M2
		VAT/SAT ratio	As provided.	M1
		CHG VAT/SAT ratio	Calculated as: postbaseline VAT/SAT ratio – baseline VAT/SAT ratio	M2
		VAT target value of <100 cm^2	Response = yes if the postbaseline VAT value <100 cm^2	Missing if postbaseline value is missing

Abbreviations: %CHG = Percent change from baseline; BMI = body mass index; CHG = Change from baseline; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; CRF = case report form; CSV = comma separated values; CT = computerized tomography; DBP = diastolic blood pressure; eGFR = estimated glomerular filtration rate; FFA = free fatty acids; FG = fasting glucose; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; HOMA = homeostasis model assessment; HOMA2 B = homeostasis model assessment 2 estimates steady state beta cell function; HOMA-S = homeostasis model assessment estimates steady state insulin sensitivity; hsCRP = high-sensitivity C-reactive protein; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite Clinical Trials; JASSO = Japan Society for the Study of Obesity; JP = Japan; JSN = Japanese Society of Nephrology; LDL = low-density lipoprotein; MCS = Mental Component Score; PCS = Physical Component Score; PGI-C= Patient Global impression of Change; PGI-S = Patient Global Impression of Severity; PROMIS = Patient-Reported Outcomes Measurement Information System; SAT = subcutaneous adipose tissue, SBP = systolic blood pressure; SD = standard deviation; SF-36 = Short Form 36-item Health Survey; SMBG = self-monitored blood glucose; T2D = type 2 diabetes; UACR = urine albumin-creatinine ratio; VAS = visual analog scale; VAT = visceral adipose tissue, VLDL = very low-density lipoprotein.

[a] M1 = Single item, missing if missing; M2 = Missing if baseline or postbaseline value is missing.

[b] Derive eGFR using the calculator at https://jsn.or.jp/medic/guideline/pdf/guide/viewer.html?file=001-294.pdf&page=39

[c] Derive HOMA2 B and HOMA2 S using the calculator at https://www.rdm.ox.ac.uk/about/our-clinical-facilities-and-mrc-units/DTU/software/homa

[d] Derive EQ-5D-5L JP population-based index score using the calculator at https://www.niph.go.jp/journal/data/64-1/201564010008.pdf

[e] Scoring algorithm obtained from the developer, Northwestern University (Chapman 2022).

Table GZPD.4.6. Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Estimand Data Points Set / Population[a]	Analysis Method[bc]	Time Point[d]	Analysis Type
Body weight	%CHG in body weight (%)	TR	ANCOVA with PMI	Week 72	Primary endpoint, main analysis
		Efficacy	MMRM	Week 72	Primary endpoint, supplementary analysis
		TR / (T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy / (T2D at baseline)	MMRM	Week 72	Exploratory analysis
		TR / (non-T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy (Non-T2D at baseline)	MMRM	Week 72	Exploratory analysis
		TR	ANCOVA with MI-TP	Week 72	Primary endpoint, sensitivity analysis
		TR	CDF plot with PMI	Week 72	Exploratory analysis
		Efficacy	CDF plot with MI	Week 72	Exploratory analysis
	Percentage of participants with ≥5% body weight reduction from baseline	TR	Logistic regression with PMI	Week 72	Primary endpoint, main analysis
		Efficacy	Logistic regression with MI	Week 72	Primary endpoint, supplementary analysis
		TR	Logistic regression with MI-TP	Week 72	Primary endpoint, sensitivity analysis
	Percentage of participants with ≥x% body weight reduction from baseline, where x = 7, 10, 15, 20	TR	Logistic regression with PMI	Week 72	Secondary endpoint
		Efficacy	Logistic regression with MI	Week 72	Secondary endpoint
	Percentage of participants with ≥x% body weight reduction from baseline,	TR / (T2D at baseline)	Logistic regression with PMI	Week 72	Exploratory analysis
		Efficacy / (T2D at baseline)	Logistic regression with MI	Week 72	Exploratory analysis

	where x = 5, 7, 10, 15, 20	TR / (non-T2D at baseline)	Logistic regression with PMI	Week 72	Exploratory analysis
		Efficacy / (non-T2D at baseline)	Logistic regression with MI	Week 72	Exploratory analysis
	CHG in body weight (kg)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy	MMRM	Week 72	Secondary endpoint
	Time to initially reach achieve body weight reduction of >= x %, where x = 5, 10, 15, 20	Efficacy	Kaplan-Meier		Exploratory analysis
BMI (remove BMI category from model)	CHG in BMI (kg/m^2)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy	MMRM	Week 72	Secondary endpoint
		TR / (non-T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy / (non-T2D at baseline)	MMRM	Week 72	Exploratory analysis
	Percentage of participants with BMI target value of x, where x = <27 kg/m^2 ,<25 kg/m^2	TR	Logistic regression with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	Logistic regression with MI (pooled OFG doses)	Week 72	Secondary endpoint
		TR / (non-T2D at baseline)	Logistic regression with PMI (pooled OFG doses)	Week 72	Exploratory analysis
		Efficacy / (non-T2D at baseline)	Logistic regression with MI (pooled OFG doses)	Week 72	Exploratory analysis
Waist circumference	CHG in waist circumference (cm)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy	MMRM	Week 72	Secondary endpoint
		TR / (T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy / (T2D at baseline)	MMRM	Week 72	Exploratory analysis
		TR / (non-T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy / (non-T2D at baseline)	MMRM	Week 72	Exploratory analysis
Blood pressure	CHG in SBP (mmHg)	TR	ANCOVA with PMI	Week 72	Secondary endpoint

		Efficacy	MMRM	Week 72	Secondary endpoint
	CHG in DBP (mmHg)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy	MMRM	Week 72	Secondary endpoint
	CHG in SBP (mmHg)	TR / (Hypertension at baseline)	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy / (Hypertension at baseline)	MMRM	Week 72	Secondary endpoint
	CHG in DBP (mmHg)	TR / (Hypertension at baseline)	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy / (Hypertension at baseline)	MMRM	Week 72	Secondary endpoint
	CHG in SBP (mmHg)	TR / (T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy / (T2D at baseline)	MMRM	Week 72	Exploratory analysis
	CHG in DBP (mmHg)	TR / (T2D at baseline)	ANCOVA with PMI	Week 72	Exploratory analysis
		Efficacy / (T2D at baseline)	MMRM	Week 72	Exploratory analysis
Lipid parameter	%CHG in total cholesterol (%)	TR / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	ANCOVA with PMI (log transformation)	Week 72	Secondary endpoint
		Efficacy / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	MMRM (log transformation)	Week 72	Secondary endpoint
	%CHG in non-HDL-cholesterol (%)	TR / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	ANCOVA with PMI (log transformation)	Week 72	Secondary endpoint
		Efficacy / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	MMRM (log transformation)	Week 72	Secondary endpoint

	%CHG in triglycerides (%)	TR / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	ANCOVA with PMI (log transformation)	Week 72	Secondary endpoint
		Efficacy / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	MMRM (log transformation)	Week 72	Secondary endpoint
	%CHG in LDL-cholesterol (%)	TR / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	ANCOVA with PMI (log transformation)	Week 72	Secondary endpoint
		Efficacy / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	MMRM (log transformation)	Week 72	Secondary endpoint
	%CHG in HDL-cholesterol (%)	TR / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	ANCOVA with PMI (log transformation)	Week 72	Secondary endpoint
		Efficacy / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	MMRM (log transformation)	Week 72	Secondary endpoint
	%CHG in VLDL-cholesterol (%)	TR / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	ANCOVA with PMI (log transformation)	Week 72	Secondary endpoint
		Efficacy / (Overall), (Dyslipidemia at baseline), (T2D at baseline)	MMRM (log transformation)	Week 72	Secondary endpoint
Free fatty acids	CHG in FFAs (mmol/L, mEq/L)	Efficacy	MMRM	Week 72	Secondary endpoint
	%CHG in FFAs (%)	Efficacy	MMRM	Week 72	Secondary endpoint

Glycemic control	CHG in fasting glucose (mg/dL and mmol/L)	Efficacy / (Overall), (T2D at baseline), (non-T2D at baseline)	MMRM	Week 72	Secondary endpoint, exploratory endpoint (non-T2D at baseline)
	CHG in fasting glucose (mg/dL and mmol/L)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
	CHG in HbA1c (% and mmol/mol)	Efficacy / (Overall), (T2D at baseline), (non-T2D at baseline)	MMRM	Week 72	Secondary endpoint, exploratory endpoint (non-T2D at baseline)
		TR / (Overall), (T2D at baseline), (non-T2D at baseline)	ANCOVA with PMI	Week 72	Secondary endpoint, exploratory endpoint (non-T2D at baseline)
	%CHG in Fasting C-peptide (mg/dL)	Efficacy	MMRM	Week 72	Secondary endpoint
	CHG in mean of all 7-point SMBG (mg/dL)	Efficacy / (only T2D at baseline)	MMRM	Week 72	Secondary endpoint
	CHG in point x SMBG (mg/dL), where x = 1, …, 7	Efficacy / (only T2D at baseline)	MMRM	Week 72	Secondary endpoint
	CHG in mean of pre-meal SMBG (mg/dL)	Efficacy / (only T2D at baseline)	MMRM	Week 72	Secondary endpoint
	CHG in post-meal SMBG (mg/dL)	Efficacy / (only T2D at baseline)	MMRM	Week 72	Secondary endpoint
	Time to onset of T2D	TR / (only nonT2D at baseline)	Cox-proportional hazards model (pooled OFG doses with baseline FSG as covariate and sex as factor)	Week 72	Secondary endpoint
Insulin sensitivity	%CHG in fasting insulin (%)	Efficacy	MMRM (log transformation)	Week 72	Secondary endpoint
Renal function	%CHG in UACR (%)	Efficacy	MMRM (log transformation) (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in eGFR JSN Cystatin-C (ml/min/1.73m^2)	Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in eGFR JSN Creatinine (ml/min/1.73m^2)	Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
Inflammatory biomarker	CHG in hsCRP (mg/L, nmol/L)	Efficacy	MRMM	Week 72	Secondary endpoint
	%CHG in hsCRP (%)	Efficacy	MRMM	Week 72	Secondary endpoint

Biomarker	CHG in uric acid (mg/dL)	Efficacy	MRMM	Week 72	Secondary endpoint
	CHG in adiponectin (µg/dL) (both high molecular weight and total adiponectin)	Efficacy	MRMM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in leptin (ng/dL)	Efficacy	MRMM (pooled OFG doses)	Week 72	Secondary endpoint
HOMA	CHG in HOMA2 B (insulin)	Efficacy / (only T2D at baseline)	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in HOMA2 S (insulin)	Efficacy / (only T2D at baseline)	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in HOMA2 B (C-peptide)	Efficacy / (only T2D at baseline)	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in HOMA2 S (C-peptide)	Efficacy / (only T2D at baseline)	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
SF-36	CHG in SF-36v2 acute form Physical Functioning domain score	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in SF-36v2 acute form domain scores (all except Physical Functioning), PCS and MCS	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
PGI-S	PGI-S rating	TR	Shift from baseline (pooled OFG doses)	Week 24, 72	Secondary endpoint
		Efficacy	Shift from baseline (pooled OFG doses)	Week 24, 72	Secondary endpoint
PGI-C	PGI-C rating	TR	Descriptive statistics (pooled OFG doses)	Week 24, 72	Secondary endpoint
		Efficacy	Descriptive statistics (pooled OFG doses)	Week 24, 72	Secondary endpoint
IWQOL-Lite-CT	CHG in physical function composite score	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in physical domain score	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in psychosocial domain score	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint

		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
	CHG in total score	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
EQ-5D-5L	CHG in EQ-5D-5L utility index and VAS	TR	ANCOVA with PMI	Week 72	Secondary endpoint
		Efficacy	MMRM	Week 72	Secondary endpoint
PROMIS	CHG in total score	TR	ANCOVA with PMI (pooled OFG doses)	Week 72	Secondary endpoint
		Efficacy	MMRM (pooled OFG doses)	Week 72	Secondary endpoint
Improvement of comorbidities	Hypertension improvement	TR / (only for participants with hypertension at baseline)	Logistic regression with observed value (pooled OFG doses)	Week 72	Secondary endpoint
	LDL-cholesterol category (≥140, <140 mg/dL)	TR / (only for participants with dyslipidemia at baseline)	Shift from baseline with PMI (pooled OFG doses)	Week 72	Additional secondary
	TG category (≥150, < 150 mg/dL (1.7 mmol/L))	TR / (only for participants with dyslipidemia at baseline)	Shift from baseline with PMI (pooled OFG doses)	Week 72	Additional secondary
	HDL-cholesterol category (<40, ≥40 mg/dL (1.0 mmol/L))	TR / (only for participants with dyslipidemia at baseline)	Shift from baseline with PMI (pooled OFG doses)	Week 72	Additional secondary
	Dyslipidemia improvement	TR / (only for participants with dyslipidemia at baseline)	Logistic regression with observed value (pooled OFG doses)	Week 72	Secondary endpoint
	T2D (HbA1c target value of <6.5%)	TR / (only T2D at baseline)	Logistic regression with PMI (pooled OFG doses)	Week 72	Secondary endpoint
	HbA1c target value of <5.7%	TR / (only T2D at baseline)	Logistic regression with PMI (pooled OFG doses)	Week 72	Secondary endpoint
	HbA1c target value of <7%	TR / (only T2D at baseline)	Logistic regression with PMI (pooled OFG doses)	Week 72	Secondary endpoint
	T2D (HbA1c target value of <6.5%)	Efficacy / (only T2D at baseline)	Logistic regression with MI (pooled OFG doses)	Week 72	Secondary endpoint
	HbA1c target value of <5.7%	Efficacy / (only T2D at baseline)	Logistic regression with MI (pooled OFG doses)	Week 72	Secondary endpoint

	HbA1c target value of <7%	Efficacy / (only T2D at baseline)	Logistic regression with MI (pooled OFG doses)	Week 72	Secondary endpoint
	Time to initially reach HbA1c <6.5%	Efficacy / (only T2D at baseline)	Kaplan – Meier (pooled OFG doses)		exploratory
	Time to initially reach HbA1c <5.7%	Efficacy / (only T2D at baseline)	Kaplan – Meier (pooled OFG doses)		exploratory
	Time to initially reach HbA1c <7%	Efficacy / (only T2D at baseline)	Kaplan – Meier (pooled OFG doses)		exploratory
	Change in glycemic category	TR	Shift from baseline (pooled OFG doses)	Week 72	Secondary endpoint
	Improvements of obesity-related health problems (0, 1)	TR / (only for participants with 1 obesity-related health problems)	Fisher's exact with observed value (pooled OFG doses)	Week 72	Secondary endpoint
	Improvements of obesity-related health problems (0, 1, 2)	TR / (only for participants with 2 obesity-related health problems)	Fisher's exact with observed value (pooled OFG doses)	Week 72	Secondary endpoint
	Improvements of obesity-related health problems (0, 1, 2, 3)	TR / (only for participants with 3 obesity-related health problems)	Fisher's exact with observed value (pooled OFG doses)	Week 72	Secondary endpoint
	Change in anti-hypertensive medication	TR / only hypertension at baseline	Descriptive statistics (pooled OFG doses)	Week 72	Secondary endpoint
	Change in lipid-lowering medication	TR / only dyslipidemia at baseline	Descriptive statistics (pooled OFG doses)	Week 72	Secondary endpoint
	Change in concomitant oral anti-glycemic medication	TR / only T2D at baseline	Descriptive statistics (pooled OFG doses)	Week 72	Secondary endpoint
VAT, SAT	CHG VAT (cm^2)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
	CHG SAT (cm^2)	TR	ANCOVA with PMI	Week 72	Secondary endpoint
	CHG VAT/SAT ratio	TR	ANCOVA with PMI	Week 72	Secondary endpoint
	VAT target value of <100 cm^2	TR /	Logistic regression with PMI	Week 72	Secondary endpoint

		(only for participants with VAT ≥100 cm^2 at baseline)			

Abbreviations: %CHG = Percent change from baseline; ANCOVA = analysis of covariance; BMI = body mass index; CDF = cumulative distribution function; CHG = Change from baseline; DBP = diastolic blood pressure; eGFR = estimated glomerular filtration rate; FFA = free fatty acids; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; HOMA = homeostasis model assessment; HOMA2 B = homeostasis model assessment 2 estimates steady state beta cell function; HOMA2 S = homeostasis model assessment estimates steady state insulin sensitivity; hsCRP = high-sensitivity C-reactive protein; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite Clinical Trials; JSN = Japanese Society of Nephrology; LDL = low-density lipoprotein; MCS = Mental Component Score; MI = multiple imputation; MI-TP = multiple imputation-based tipping-point; MMRM = mixed-effects model for repeated measures; OFG = orforglipron; PCS = Physical Component Score; PGI-C= patient global impression of change; PGI-S = patient global impression of severity; PMI = primary multiple imputation; PROMIS = Patient-Reported Outcomes Measurement Information System; SAT = subcutaneous adipose tissue ; SBP = systolic blood pressure; SF-36 = Short Form 36-item Health Survey; SMBG = self-monitored blood glucose; T2D = type 2 diabetes; TG = triglycerides; TP = tipping point; TR = Treatment Regimen; UACR = urine albumin-creatinine ratio; VAS = visual analog scale; VAT = visceral adipose tissue; VLDL = very low-density lipoprotein.

[a] Population is displayed if different from "Randomized Participants."

[b] PMI strategy will be performed after the log transformation.

[c] Pooled OFG doses: In addition to analyzing each OFG dose, a pooled analysis for the OFG dose is conducted.

[d] Assessments collected at multiple postbaseline visits will be analyzed at those scheduled visits using MMRM as supplemental analyses, in addition to the primary time point listed in the table.

4.4.1. Secondary Endpoints

PK/PD analyses will be documented in a separate document.

4.4.1.1. Definition of Endpoint(s)

The definitions of the secondary endpoints are specified in Table GZPD.4.5.

4.4.1.2. Main Analytical Approach

The analytical approaches for the secondary endpoints are specified in Table GZPD.4.6.

4.4.1.3. Sensitivity Analyses

No sensitivity analyses are planned for the secondary endpoints.

4.4.1.4. Supplementary Analyses

No supplementary analyses are planned for the secondary endpoints.

4.4.2. Additional Secondary Endpoints

The definitions of additional secondary endpoints are specified in Table GZPD.4.5.

Additionally, the analytical approaches are specified in Table GZPD.4.6.

4.5. Exploratory Endpoints Analysis

PK/PD analyses will be documented in a separate document.

4.6. Safety Analyses

The planned safety analyses are consistent with compound-level safety standards, which are based on various sources, including company standards, internal and external subject matter experts, publications from cross-industry initiatives (for example, PHUSE 2013, 2015, 2017, 2018, 2022), and publications from regulatory agencies (for example, EMA 2014, CDER/BIRRS 2022a, 2022b). Descriptions of the safety analyses are provided in this SAP, but some details are found in the compound-level safety standards.

For the interpretation of safety analysis results, p-values will not be used for hypothesis testing. For the purposes of these analyses, p-values will be considered as a number between 0 and 1 that give an idea of how strong the evidence is for an imbalance between study intervention arms. The evidence for an imbalance is stronger when the value is toward 0 and weaker when the value is toward 1. Similarly, CIs will not be used for hypothesis testing. They reflect the uncertainty of an estimate.

Unless otherwise specified, safety analyses will be conducted using the safety participants set (Table GZPD.3.1) and the safety data points set (Table GZPD.3.2). It is noted that for some safety endpoints, such as treatment discontinuation due to AE, and so on, the analyses will be based on data observed during treatment period only.

Summary tables with risk difference will be sorted by decreasing order of risk difference.

4.6.1. Extent of Exposure

Duration of exposure to study treatment will be summarized by treatment group for safety participants. Descriptive statistics for participant weeks (or days) of exposure and total participant years in exposure will be provided. Overall exposure will be summarized in total participant-year (PY) of exposure, derived in the following manner:

total PY of exposure = sum of duration of exposure in days (for all participants in treatment group) / 365.25

The frequency of participants falling into different exposure ranges will be summarized:

- >0
- ≥4 weeks
- ≥8 weeks
- ≥12 weeks
- ≥16 weeks
- ≥20 weeks
- ≥24 weeks
- ≥36 weeks
- ≥48 weeks
- ≥60 weeks, and
- ≥72 weeks.

No p-values will be reported.

4.6.2. Adverse Events

The planned summaries for AEs are provided in Table GZPD.4.7 and are described more fully in the compound-level safety standards.

Table GZPD.4.7. Summary Tables Related to Adverse Events

Analysis	Method	Population/Period
Overview of AEs, including • TEAE • SAE • death, and • permanent discontinuation from study intervention due to an AE.	Fisher's exact	Safety/TP + FP (TP for study treatment discontinuation due to an AE)
TEAEs by PT within SOC	Fisher's exact	Safety/TP + FP
TEAEs by PT	Fisher's exact	Safety/TP + FP
Maximum Severity TEAEs by PT within SOC		Safety/TP + FP
TEAEs with incidence ≥5% by PT	Fisher's exact	Safety/TP + FP
SAEs by PT within SOC	Fisher's exact	Safety/TP + FP
Primary AEs leading to permanent discontinuation of study intervention by PT within SOC	Fisher's exact	Safety/TP
Primary AEs leading to permanent discontinuation of study by PT within SOC	Fisher's exact	Safety/TP
AEs leading to study intervention interruption by PT within SOC	Fisher's exact	Safety/TP
AEs leading to study intervention modifications by PT within SOC	Fisher's exact	Safety/TP
Listing of SAEs		Safety/TP + FP
Listing of primary AEs leading to permanent discontinuation of study intervention		Safety/TP
Listing of primary AEs leading to permanent discontinuation of study		Safety/TP + FP
Listing of deaths		Safety/TP + FP
Listing AEs related to suspected overdosing		Safety/TP
Narratives for participants with at least 1 notable event		Randomized/TP + FP

Abbreviations: AE=adverse event; FP = follow-up period; PT = preferred term; SAE = serious adverse event; SOC = system organ class; TEAE = treatment emergent adverse event; TP = treatment period.

4.6.3. Clinical Laboratory Evaluation

The planned summaries for clinical laboratory evaluations are provided in Table GZPD.4.8 and are described more fully in the compound-level safety standards.

Table GZPD.4.8. Summary Tables Related to Clinical Laboratory Evaluations.

Analysis	Method	Population/Period
Box plots and mean/SD for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plots and mean/SD for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Summary for participants with elevated or low values meeting specified levels	Descriptive statistics	Safety/TP + FP
Listing of abnormal laboratory findings		Safety/TP + FP

Abbreviations: FP = follow-up period; SD = standard deviation; TP = treatment period.

4.6.4. Vital Signs and Physical Characteristics

Triplicate vital signs will be collected at each visit, thus the mean of these measurements will be used for the vital signs analyses. The planned summaries for vital signs (systolic blood pressure,

diastolic blood pressure, and pulse rate) are provided in Table GZPD.4.9, and are described more fully in the compound-level safety standards.

Table GZPD.4.9. Summary Tables Related to Vital Signs

Analysis	Method	Population/Period
Box plots and mean/SD for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plots and mean/SD for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Analysis of blood pressure and pulse rate for change from baseline	MMRM	Safety/TP + FP
Summary for participants meeting specific blood pressure and pulse rate levels	Descriptive statistics	Safety/TP + FP
Shift of maximum-to-maximum for pulse rate	Descriptive statistics	Safety/TP + FP

Abbreviations: FP = follow-up period; MMRM = mixed model for repeated measures; SD = standard deviation; TP = treatment period.

4.6.5. Electrocardiograms

Triplicate electrocardiograms (ECGs) will be collected at the same visit; thus the mean of these measurements will be used for the analysis. The planned summaries for ECG parameters are provided in Table GZPD.4.10 and are described more fully in the compound-level safety standards.

Table GZPD.4.10. Summary Tables Related to ECG Parameters

Analysis	Method	Population/Period
Box plot and mean/SD for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plot and mean/SD for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Heart rate and PR interval in specified categories	Descriptive statistics	Safety/TP + FP
Analysis for change from baseline in ECG parameters (heart rate, PR interval)	MMRM	Safety/TP + FP

Abbreviations: ECG=electrocardiogram ; FP = follow-up period; MMRM = mixed model for repeated measures; PR = pulse rate; SD = standard deviation; TP = treatment period.

4.6.6. Safety Topics of Interest

This section includes safety topics of interest whether due to observed safety findings, potential findings based on drug class, or safety topics anticipated to be requested by a regulatory agency for any reason. In general, safety topics of interest will be identified by one or more standardized Medical Dictionary for Regulatory Activities (MedDRA) query(ies) (SMQs), system organ class (SOC), high-level term (HLT), FDA Medical Query (FMQ), or a Lilly defined MedDRA preferred term (PT) listing based upon the review of the most current MedDRA Version, or by relevant laboratory changes. Search criteria are detailed in the compound-level safety standards.

The planned analyses for safety topics of interest are provided in Table GZPD.4.11 and are described more fully in the compound-level safety standards.

Table GZPD.4.11. Description and Analyses of Safety Topics of Interest

Special Safety Topic	Short Description	Analysis	Method	Population/Period
Major Adverse Cardiovascular Events	Death and nonfatal CV AEs will be adjudicated by a committee of physicians external to Lilly with cardiology expertise: a CEC.	Positively adjudicated MACE by category/subcategory MACE reported by investigators may be summarized		Safety/TP + FP
		Listing of all MACE reported by investigator (whether or not positively adjudicated)		Safety/TP + FP
Arrhythmias and Cardiac Conduction Disorders	The TE arrhythmias and cardiac conduction disorders events will be derived using the MedDRA PTs contained in certain SMQs and HLT.	TE arrhythmias and cardiac conduction disorders by PT nested within SMQ and HLT[a].	Fisher's exact	Safety/TP + FP
Hypotension, Orthostatic Hypotension, and Syncope	The TE hypotension, orthostatic hypotension, and syncope events will be derived using MedDRA PTs and FMQs.	TE hypotension, orthostatic hypotension, and syncope by PT[a].	Fisher's exact	Safety/TP + FP
		Hypotension by baseline antihypertensive medication use using hypotension narrow FMQ		
Hypoglycemia[ab]	The 2023 American Diabetes Association position statement on glycemic targets (El Sayed et al. 2023) will be used to define: • Level 1 hypoglycemia • Level 2 hypoglycemia • Level 3 hypoglycemia (severe/serious hypoglycemia) Nocturnal hypoglycemia events (including severe hypoglycemia) **occur at night** and presumably during sleep.	Incidence (percent of patients experiencing 1 or more episode) of level 2 or level 3 hypoglycemia.	Logistic regression	Safety/TP + FP excluding events after initiation of new antihyperglycemic therapy as defined in Section 6.3 and Table GZPD.6.4. Supportive analysis: Safety/TP + FP regardless of initiation of new antihyperglycemic therapy
		Incidence of level 3 (severe/serious) hypoglycemia		
		Incidence of level 1 hypoglycemia (if warranted by data)[a].		
		Rate (episodes/patient/year) of level 2 or level 3 hypoglycemia.	Negative binomial regression (the logarithm of days during the analysis interval will be adjusted as an offset)	
		Rate of level 3 (severe/serious) hypoglycemia.		
		Rate of level 1 hypoglycemia (if warranted by data).		
		Listing of level 2 or level 3 hypoglycemia events.		
		Listing of level 3 or level 3 nocturnal hypoglycemia events.		

Severe Persistent Hyperglycemia	Initiation of new antihyperglycemic therapy taken for severe persistent hyperglycemia	Listing and Summary of Rescue therapy taken for severe persistent hyperglycemia(if warranted by data)[a].	Fisher's exact	Safety with T2D at baseline/TP + FP
Gastrointestinal (GI) Safety[c]	GI AEs using Gastrointestinal disorders SOC and FMQ will be captured.	Severe or serious TE GI events by PT.	Fisher's exact	Safety/TP + FP
		Study intervention discontinuation due to TE GI events.	Fisher's exact	Safety/TP + FP
		TE abdominal pain by FMQ narrow PT	Fisher's exact	Safety/TP + FP
		TE nausea, vomiting, diarrhea, constipation, and NVD by maximum severity.	Fisher's exact	Safety/TP + FP
		Prevalence and incidence over time for TE nausea, vomiting, diarrhoea, constipation, and NVD.		Safety/TP + FP
		Plot of time to the onset of TE nausea, vomiting, diarrhoea, constipation, and NVD.	Kaplan-Meier	Safety/TP + FP
		Plot of prevalence and incidence over time for TE nausea, vomiting, diarrhoea, constipation, and NVD by maximum severity.		Safety/TP + FP
Renal Safety	Laboratory measures related to renal safety will be analyzed. Renal events including acute renal failure and chronic renal failure exacerbation will be captured using SMQs. Dehydration events will be captured using SMQs.	Shift of minimum-to-minimum for eGFR estimated by the JSN equation[d].		Safety/TP + FP
		Shift of maximum-to-maximum for UACR.		Safety/TP + FP
		Summary of eGFR decrease from baseline		Safety/TP + FP
		MMRM analyses for eGFR.	MMRM	Safety/TP + FP
		MMRM analyses for UACR (log transformation).	MMRM	Safety/TP + FP
		TE renal events by PT nested within SMQ[a].	Fisher's exact	Safety/TP + FP
		Listing of TE dehydration events by PT[a].		Safety/TP + FP
Pancreatic Safety	The pancreatic enzyme data (p-amylase and lipase) will be observed through laboratory testing. All suspected cases of acute or chronic pancreatitis and AEs of severe or serious abdominal pain of unknown etiology will be sent for adjudication by an independent CEC.	Shift of maximum-to-maximum for pancreatic enzyme		Safety/TP + FP
		MMRM analysis for pancreatic enzymes (p-amylase and lipase) with a log transformation (postbaseline/baseline).	MMRM	Safety/TP + FP
		Investigator-reported events and subsequently confirmed adjudicated events, respectively	Fisher's exact	Safety/TP + FP
		TE Pancreatic events by "Acute pancreatitis" SMQ (narrow scope) and "Chronic pancreatitis" PT		Safety/TP + FP

		Listing of adjudicated and investigator-reported pancreatitis.		Safety/TP + FP
Thyroid Malignancies and C-Cell Hyperplasia	TE thyroid malignancies and C-cell hyperplasia will be identified using MedDRA HLT and PT. The purpose of calcitonin measurements is to assess the potential effect of orforglipron on thyroid C-cell function.	TE thyroid C-cell hyperplasia and malignancies by PT.	Fisher's exact	Safety/TP + FP
		Summary of maximum postbaseline calcitonin value in the thresholds.		Safety/TP + FP
		• Listing of participants who meet protocol defined discontinuation criteria based on calcitonin: with eGFR <60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥35 ng/L AND ≥50% increase from the baseline value[a] • with eGFR ≥60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥20 and <35 ng/L AND ≥50% increase from the baseline value and ≥10% increase on retest[a] with eGFR ≥60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥35 ng/L AND ≥50% increase from the baseline value[a]		Safety/TP + FP
		MMRM analysis for calcitonin (log transformation).	MMRM	Safety/TP + FP
Malignancies	The malignancy events will be derived using the MedDRA PTs containing certain SMQs.	TE malignancy by PT nested within SMQ[a].	Fisher's exact	Safety/TP + FP
Hepatic Safety	Hepatic labs include ALT, AST, ALP, TBL, DBL, and GGT. When criteria are met for hepatic evaluations, investigators will conduct close monitoring of hepatic symptoms and liver tests, perform a comprehensive evaluation for alternative causes of abnormal liver tests, and complete follow-up hepatic safety eCRFs. Potentially drug-related hepatic disorders will be identified using SMQs.	Abnormal postbaseline categories for hepatic safety parameters: ALT, AST, ALP, TBL, DBL, and GGT.		Safety/TP + FP
		Treatment-emergent potentially drug-related hepatic disorders by PT nested within SMQ		Safety/TP + FP
		Hepatocellular drug-induced liver injury screening plot (maximum TBL vs maximum ALT or AST).		Safety/TP + FP
		Hepatocellular drug-induced liver injury screening table.		Safety/TP + FP
		Cholestatic drug-induced liver injury screening table		Safety/TP + FP

		Cholestatic drug-induced liver injury screening plot (maximum TBL vs maximum ALP).		Safety/TP + FP
		Hepatic Laboratory Parameters (ALT, AST, ALP, TBL, DBL, and GGT) with a log transformation (postbaseline/baseline)	MMRM	Safety/TP + FP
		Listing participants with ALT or AST ≥3X ULN		Safety/TP + FP
		Listing of participants with ALP or TBL ≥2X ULN		Safety/TP + FP
		Shift of maximum-to-maximum for ALT, AST, ALP, TBL		
		Participant profiles for participants meeting criteria for a comprehensive hepatic evaluation (as defined in the protocol).		Safety/TP + FP
Gallbladder and Biliary Tract Disorders	All events of TE gallbladder and biliary tract disorders will be identified by using certain SMQs.	TE gallbladder and biliary tract disorders by PT nested within SMQ[a].	Fisher's exact	Safety/TP + FP
Hypersensitivity Reactions	All events of TE allergic reaction and hypersensitivities will be identified by using certain SMQs.	TE allergic reaction and hypersensitivities by PT nested within SMQ[a].	Fisher's exact	Safety/TP + FP
Depression, Disorder/ Suicidal Ideation and Behavior	AEs will be searched using MedDRA PTs that satisfies the search criteria.	TE major depressive disorder, suicidal ideation, or behavior events by PT nested within applicable FMQ or SMQ[a].	Fisher's exact	Safety/TP + FP
	Suicide-related thoughts and behaviors will be collected based on the C-SSRS.	Summary of C-SSRS categories and composite measures.		Safety/TP + FP
		Listing of C-SSRS categories and composite measures		Safety/TP + FP
	PHQ-9 will be collected to assesses the specific diagnostic symptoms that determine the presence of a clinical depressive disorder. The PHQ-9 total scores will be categorized as none (not depressed), mild, moderate, moderately severe, and severe.	Shift of each baseline category (maximum value) versus each postbaseline category (maximum value)		Safety/TP + FP
		Summary of categories based on the maximum values during baseline and postbaseline: • any increase in depression category (that is, worsening of depression) • increase from Minimal to none or Mild depression to Moderate, Moderately severe, or Severe depression, and		Safety/TP + FP

		• increase from Mild or Moderate depression to Moderately severe or Severe depression.		
Abuse Potential	AEs will be searched using a modified abuse potential FMQ.	TE abuse potential events by PT[a].	Fisher's exact	Safety/TP + FP
AEs possibly related to loss of lean muscle mass	AEs will be identified using MedDRA HLTs and PTs	TEAEs possibly related to loss of lean muscle mass by PT within Event Category	Fisher's exact	Safety/TP + FP
Malnutrition	AEs will be identified using MedDRA HLTs and PTs	TE malnutrition by PT within Event Category	Fisher's exact	Safety/ TP +FP
Excessive weight loss	AEs will be identified using MedDRA PTs BMI< 18.5 kg/m^2 will be analyzed	Severe or serious TE excessive weight loss by PT or BMI <18.5 kg/m^2	Fisher's exact	Safety/ TP + FP
		Listing of Participants with BMI< 18.5 kg/m^2		Safety/ TP + FP
Diabetic Retinopathy	Diabetic retinopathy and macular edema will be monitored using retinal fundus photography and additional ophthalmological follow-up as warranted throughout the study. Diabetic retinopathy, macular edema, and related complication will be searched using MedDRA PTs.	Baseline and postbaseline fundus photography exam by treatment group and by eye (that is, any eye, left eye and right eye).		Safety with T2D at baseline/TP + FP
		Shift of baseline diabetic retinopathy and macular edema status to most severe postbaseline diabetic retinopathy and macular edema status.		Safety with T2D at baseline//TP + FP
		TE diabetic retinopathy complications by PT.	Fisher's exact	Safety with T2D at baseline//TP + FP
		Listing of Treatment-Emergent Diabetic Retinopathy, Diabetic Macular Edema, and Related Complications		Safety with T2D at baseline//TP + FP

Abbreviations: AE = adverse event; ALP = serum alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; CEC = clinical endpoint committee; CKD-EPI = Chronic Kidney Disease Epidemiology; C-SSRS = Columbia suicide severity rating scale; CV = cardiovascular; DBL = direct bilirubin; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; FDA = United States Food and Drug Administration; FMQ = FDA Medical Query; FP = follow-up period; GGT = gamma-glutamyl transferase; GI = gastrointestinal; HLT = high-level term; INR = international normalized ratio; JSN = Japanese Society of Nephrology; MACE = major adverse cardiovascular event; MedDRA = Medical Dictionary for Regulatory Activities; MMRM = mixed model for repeated measures; NVD = nausea, vomiting, diarrhea; PHQ-9 = Patient health questionnaire-9; PT = preferred term; SMQ = standardized MedDRA query; SOC = system organ class; T2D = type 2 diabetes; TBL = total bilirubin; TE = treatment-emergent; TP = treatment period; UACR = urinary albumin-to-creatinine ratio; vs = versus.

[a] For these tables, if the number of events is less than 10, a listing will be provided instead.

[b] The participants with T2D and those without T2D are distinguished and analyzed separately at baseline.

[c] Additionally, all GI events will be analyzed.

[d] Derived using JSN eGFRcr formula at https://jsn.or.jp/medic/guideline/pdf/guide/viewer.html?file=001-294.pdf

Note: Listings and participant profiles may be provided through interactive display tools instead of a static display.

4.7. Other Analyses

4.7.1. Subgroup Analyses

Subgroup analyses will be conducted for the primary endpoints with the treatment regimen estimand:

- percent change in body weight from baseline to the Week 72 visit,
- achieving 5% or more body weight reduction from baseline to the Week 72 visit.

The variables for subgroup analysis are specified in Section 6.1 and Table GZPD.6.1.

The ANCOVA model specified in Section 4.1.2.1.1 will be fitted separately within each subgroup for percent change in body weight from baseline to the Week 72 visit. The least squares (LS) mean, LS mean difference, standard error (SE), and 95% CI will be presented. The LS means and variance-covariance estimates from these separate models will be used to test the treatment-by-subgroup interaction at a significance level of 0.10.

The logistic regression model specified in Section 4.1.2.1.4 will be fitted separately within each subgroup for the binary variable achieving 5% or more body weight reduction at the Week 72 visit. The odds ratio and 95% CI will be presented. The LS means and variance-covariance estimates from these separate models will be used to test the treatment-by-subgroup interaction at a significance level of 0.10.

If the subgroup factor is part of the strata variable, a new strata variable that excludes the subgroup categories will be used.

If any category within the subgroup is <5% of the total population, only descriptive statistics will be provided for that category (that is, there will be no inferential testing within the subgroup category).

A forest plot including the treatment difference and 95% CI estimated for each subgroup level will be presented.

A Bayesian shrinkage method will be used to provide adjusted estimates and improve estimation precision. Following the paper by Wang et. al. (2024), for a given baseline characteristic with k subgroups, let Y_i $(i = 1, \ldots, k)$ be the observed sample estimate of the treatment effect difference in subgroup i. The following hierarchical model will be used:

$$Y_i \sim N(\mu_i, \sigma_i^2),$$

$$\mu_i \sim N(\mu, \tau^2),$$

$$\mu \sim N(0, 30^2),$$

$$\tau \sim N^+(10),$$

where σ_i^2 is set to the observed variance for sample estimate, $N^+(10)$ is the half-normal distribution with a scaled parameter of 10. A weakly informative prior, as suggested by Röver et

al. (2021) and specified above, will be applied to μ and τ. Of note, the unit information standard deviation (UISD) is assumed to be 15% based on historical weight management studies, and a standard deviation of 30% is chosen for the centrality parameter μ , so that it is approximately twice the unit information standard deviation (UISD). A scale parameter of 10 is chosen for the half-normal distribution such that the median heterogeneity in percent body weight reduction is approximately 6.7%, which is a reasonably large difference.

Additional subgroup evaluations may be conducted as exploratory analyses.

4.8. Interim Analyses

No interim analyses are planned for this study. If an unplanned interim analysis is deemed necessary for reasons other than a safety concern, the protocol must be amended.

4.8.1. Data Monitoring Committee

A data monitoring committee is not planned for Study GZPD.

4.9. Changes to Protocol-Planned Analyses

Treatment compliance is defined as taking at least 75% of required study intervention during the treatment period

5. Sample Size Determination

A sample size of 236 participants (59 per treatment group) provides more than 80% power to demonstrate the superiority of 6 mg, 12 mg, and 36 mg orforglipron to placebo with regards to percent change in body weight from baseline to Week 72. The sample size determination assumes that evaluation of superiority of 6 mg, 12 mg, and 36 mg orforglipron to placebo will be conducted in parallel, each at a 2-sided significance level of 0.016, using a 2-sample t-test. Additionally, a difference of at least 7.5% mean body weight reduction from baseline at 72 weeks for 12 mg and 36 mg orforglipron compared with placebo, a common standard deviation of 10%, and a dropout rate of 20% are assumed for the statistical power calculation.

The sample size also provides more than 80% power to demonstrate the superiority of 6 mg, 12 mg, and 36 mg orforglipron to placebo in terms of percentage of participants achieving at least 5% body weight reduction at 72 weeks, conducted in parallel using a Fisher's exact test, each at a 2-sided significance level of 0.016, assuming 30% placebo-treated participants and 65% orforglipron-treated participants achieving the goal, and a dropout rate of 20%.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

Demographics and other baseline characteristics will be summarized by treatment group for randomized participants. The listing of basic demographic characteristics (that is, age, sex, ethnicity, race, country, baseline body weight, and so on) for randomized participants will be provided.

Continuous variables will be summarized using descriptive statistics. Categorical variables will be summarized using frequency counts and percentages. No inferential analysis for the comparability of baseline demographics and characteristics across treatment groups will be performed.

Table GZPD.6.1 describes specific variables and how they will be summarized. The last column specifies variables used for the subgroup analysis described in Section 4.7.1.

Table GZPD.6.1. Demographics and Baseline Characteristics with Variables for Subgroup Analysis

Variable	Quantitative Summary	Categorical Summary	Subgroup Analysis[a]
Demographics			
Age[b]	Yes	<65, ≥65 years	X
		<75, ≥75 years	
		<65, ≥65 and <75, ≥75 years	
Sex	No	Male, Female	X
Ethnicity	No	Not reported	
Race	No	Asian	
Country	No	Japan	
Height (cm)	Yes		
Baseline waist circumference (cm)	Yes		
Baseline body weight (kg)	Yes		
Baseline BMI	Yes	<30, ≥30 to <35, ≥35 to <40, ≥40 kg/m^2	
		< 35, ≥ 35 kg/m^2	X
Caffeine use	No	Never, Current, Former	
Alcohol use	No	Never, Current, Former	
Tobacco use	No	Never, Current, Former	
Nicotine Replacement use	No	Never, Current, Former	
Baseline Therapy			
Baseline number of AHMs	0, 1, 2, 3		
Baseline Disease Characteristics			
Baseline systolic blood pressure (mm Hg)	Yes		
Baseline diastolic blood pressure (mm Hg)	Yes		
Baseline pulse rate (beats/min)	Yes		
Baseline eGFR JSN Cystatin-C (mL/min/1.73m^2)	Yes	<60, <90, ≥90 mL/min/1.73m^2	
Baseline eGFR JSN Creatinine (mL/min/1.73m^2)	Yes	<60, <90, ≥90 mL/min/1.73m^2	
Baseline UACR (g/kg)	Yes	<30, ≥30 and ≤300, >300 g/kg	
Baseline hsCRP (mg/L)	Yes		
Baseline uric acid (mg/dL)	Yes		
Baseline triglycerides (mg/dL)	Yes		
Baseline total cholesterol (mg/dL)	Yes		
Baseline VLDL-cholesterol (mg/dL)	Yes		
Baseline non-HDL-cholesterol (mg/dL)	Yes		
Baseline LDL-cholesterol (mg/dL)	Yes		
Baseline HDL-cholesterol (mg/dL)	Yes		
Baseline free fatty acid (mg/dL)	Yes		
Baseline free fatty acid (mEq/L)	Yes		
Baseline fasting insulin (pmol/L)	Yes		
Baseline HbA1c (%)	Yes	<8.5% [≤69 mmol/mol], ≥8.5% [>69 mmol/mol]	
Baseline HbA1c (mmol/mol)	Yes		
Baseline fasting glucose (mg/dL)	Yes		
Baseline fasting glucose (mmol/L)	Yes		
Baseline fasting C-peptide (mg/dL)	Yes		
Duration of diagnosis for T2D (years)	Yes	≤5, >5 and ≤10, >10 years	

Duration of obesity (years)	Yes		
Impaired glucose tolerance	No	Yes, No	
Baseline obesity related health disorders[c]			
T2D	No	Yes, No	X
Glucose intolerance disorder (T2D, IGT)	No	Yes, No	
Hypertension	No	Yes, No	X
Dyslipidemia	No	Yes, No	X
Hyperuricemia and Gout	No	Yes, No	
Cardiovascular disease, myocardial infarction, and angina	No	Yes, No	
Cerebral infarction and TIA	No	Yes, No	
Menstrual disorder and infertility	No	Yes, No	
OSA and obesity-hypoventilation syndrome	No	Yes, No	
Motor dysfunction: arthritis/osteoarthritis	No	Yes, No	
NAFLD / MAFLD	No	Yes, No	
Obesity-related renal disease	No	Yes, No	
Comorbidity number category	No	None, 1, 2, 3, 4, ≥5 (Grouping may vary depend on the data)	

Abbreviations: AE = adverse event; AHM = antihyperglycemic medications; BMI = body mass index; CKD-EPI = chronic kidney disease - epidemiology; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; hsCRP = high-sensitivity C-reactive protein; IGT = impaired glucose tolerance; LDL = low-density lipoprotein; MAFLD = metabolic associated fatty liver disease; NAFLD = nonalcoholic fatty liver disease; OSA = obstructive sleep apnea; T2D = type 2 diabetes; TIA = transient ischemic attack; UACR = urine albumin-to-creatinine ratio; VLDL = very low-density lipoprotein.

[a] Subgroup analyses are defined in Section 4.7.1 with more details.

[b] Age in years will be calculated as length of the time interval from the imputed date of birth (July 1st in the year of birth collected in the eCRF) to the informed consent date.

[c] Obesity related health disorders will be based on medical history CRF.

6.2. Appendix 2: Historical Illnesses and Preexisting Conditions

Historical illness is defined as the condition or event with an end date prior to the date of informed consent.

A preexisting condition is defined as a condition or event with a start date prior to the first dose of the study intervention and a stop date that is at or after the informed consent date, or which has no stop date (that is, ongoing). The randomization date will be used if a participant has not been dosed.

The planned summaries for historical illness and preexisting conditions are provided in Table GZPD.6.2. No inferential analysis will be performed.

Table GZPD.6.2. Summary Tables Related to Historical Illness and Preexisting Conditions

Analysis	Population/Period
Historical illness by PT within SOC	Randomized
Preexisting conditions by PT within SOC	Randomized

Abbreviations: PT = preferred term; SOC = system organ class.

6.3. Appendix 3: Concomitant Medications

Medications that start before or at the last date of the treatment period or follow-up period and are ongoing, or ended during the treatment period or follow-up period, will be classified as concomitant medication. Medications that start and end before the first dose date of the study intervention will be classified as prior therapy.

Baseline is defined as the corresponding medication taken on the day before the date of the first dose of the study intervention.

If there are no doses of study intervention, the randomization date will be used instead of the first dose date.

The planned summaries for concomitant medications are provided in Table GZPD.6.3.

Table GZPD.6.3. Summary Tables Related to Concomitant Medications

Analysis	Method	Population/Period
CMs by PN	Fisher's exact	Randomized/TP + FP
Antihypertensive CMs at baseline by PN within ATC code		Randomized
Lipid lowering CMs at baseline by PN within ATC code		Randomized
Antihyperglycemic CMs at baseline by PN within ATC code		Randomized
Status change of antihypertensive, lipid lowering, and antihyperglycemic CMs		Randomized/TP
Antihyperglycemic CM initiated after baseline by PN	Fisher's exact	Randomized/TP
Antiemetic CM initiated after baseline by PN	Fisher's exact	Randomized/TP
Antidiarrheal CM initiated after baseline by PN	Fisher's exact	Randomized/TP
Prohibited concomitant medication and surgical procedure on weight management treatment by PN		Randomized Participants/TP
Prohibited concomitant medication on anti-hyperglycemic therapy by PN		Randomized Participants/TP
Obesity management medications 1 year prior to screening by PN within ATC code		Randomized

Abbreviations: ATC = Anatomical Therapeutic Chemical; CM = concomitant medication; FP = follow-up period; PN = preferred name; TP = treatment period.

Table GZPD.6.4 provides the protocol-specified concomitant medications of interest.

Table GZPD.6.4. Protocol-Specified Concomitant Medications

Category	Preferred Name / ATC Code
Antihypertensive Medication	All medications with ATC code containing C01, C03, C07, C08, or C09
Lipid Lowering Medication	All medications with ATC code containing C10
Antihyperglycemic Medication	All medications with ATC code containing A10, except for A10XA

Abbreviation: ATC = Anatomical Therapeutic Chemical.

Table GZPD.6.5 provides the protocol-specified prohibited concomitant medication and surgical procedure rules.

Table GZPD.6.5. Prohibited Concomitant Medication Rules

Timing	Preferred Name / ATC Code	ATC Classification	Rule[a]
Prohibited Concomitant Medication on Weight Management Treatment[b]			
Postbaseline	All medications with ATC code A08AA	Centrally acting antiobesity products	
	All medications with ATC code A08AB	Peripherally acting antiobesity products	
	All medications with ATC code A08AX	Other antiobesity drugs	
	LIRAGLUTIDE		
	SEMAGLUTIDE		
	TIRZEPATIDE		
	ORLISTAT		
	SIBUTRAMINE		
	PHENYLPROPANOLAMINE		
	MAZINDOL		
	PHENTERMINE		
	LORCASERIN		
	TOPIRAMATE		Any initiation is prohibited during treatment period if taken with PHENTERMINE
	BUPROPION		Any initiation is prohibited during treatment period if taken with NALTREXONE
Prohibited Surgical Procedure on Weight Management Treatment[b]			
Postbaseline	ABDOMINOPLASTY		
	GASTRIC BANDING		
	GASTRIC BYPASS		
	LIPOSUCTION		
	SLEEVE GASTRECTOMY		
	CRYOLIPOLYSIS		
	MUCOSAL ABLATION		
	GASTRIC ARTERY EMBOLIZATION		
	INTRAGASTRIC BALLOON		

	DUODENAL-JEJUNAL ENDOLUMINAL LINER		
Prohibited Concomitant Medication on Glycemic Therapy[b]			
Postbaseline	All medications with ATC code of A10BH	DPP-4 inhibitors	
	All medications with ATC code of A10BJ	GLP-1 analogues	
	DULAGLUTIDE		
	LIRAGLUTIDE		
	SEMAGLUTIDE		
	EXENATIDE		
	TIRZEPATIDE		
	PRAMLINTIDE		
	SITAGLIPTIN		
	SAXAGLIPTIN		
	LINAGLIPTIN		
	ALOGLIPTIN		
	TENELIGLIPTIN		
	BILDAGLIPTIN		

Abbreviations: ATC = Anatomical Therapeutic Chemical; DPP-4 = dipeptidyl peptidase 4; GLP-1 = glucagon-like peptide-1; ICE = intercurrent event.

[a] The rule is displayed if different from "Any initiation is prohibited during treatment period."

[b] Use of these medications will be reviewed by the study team (blinded to the study intervention) on a case-by-case basis to assess the clinical significance of the change in medication. After review, if deemed to be nonclinically significant, the case won't be considered as an ICE.

6.4. Appendix 4: Treatment Compliance

Treatment compliance is defined as taking at least 75% study intervention during the treatment period. Compliance will be calculated by

([total number of doses dispensed – total number of doses returned]/total number of doses expected to be administered) × 100%.

Frequency counts and percentages of participants compliant to the study intervention will be summarized by treatment group using the safety participants during the treatment period. No inferential analysis will be performed.

Participants with dose interruptions or modifications will be summarized with reasons. Interruptions will be summarized if the duration of the interruption is 7 days or longer for all reasons other than an AE.

6.5. Appendix 5: Important Protocol Deviations

Important protocol deviations are identified in the Trial Issues Management Plan. A listing and summary of important protocol deviations by treatment group will be provided for all randomly assigned participants. No inferential analysis will be performed.

6.6. Appendix 6: Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

Summary of AEs, provided as a dataset that will be converted to an XML file. Both serious AEs (SAEs) and 'Other' nonserious AEs are summarized by treatment group and MedDRA PT.

- An AE is considered "Serious" whether or not it is a treatment-emergent AE (TEAE).
- An AE is considered in the "Other" category if it is both a TEAE and is not serious. For each SAE and "Other" AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term, and
    - the number of events experienced.

7. References

Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7-16. https://doi.org/10.1002/pst.1548

Alosh, M, Bretz, F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med.* 2014;33(4):693-713. https://doi.org/10.1002/sim.5974

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604. https://doi.org/10.1002/sim.3495

Bretz, F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913. https://doi.org/10.1002/bimj.201000239

[CDER/BIRRS] Center for Drug Evaluation and Research and Biomedical Informatics and Regulatory Review Science Team. Advancing premarket safety analytics workshop meet material_FDA-DM FMQs. Published September 06, 2022a. Accessed January 24, 2024. https://www.regulations.gov/document/FDA-2022-N-1961-0001

[CDER/BIRRS] Center for Drug Evaluation and Research and Biomedical Informatics and Regulatory Review Science Team. Standard safety tables and figures: integrated guide. Published August, 2022b. Accessed January 24, 2024. https://www.regulations.gov/document/FDA-2022-N-1961-0046

Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Clarendon Press; 1994.

Du Y, Li J, Raha S, Qu Y. A unified Bayesian framework for bias adjustment in multiple comparisons from clinical trials. *Stat Med*. 2024;43(15):2928-2943. https://doi.org/10.1002/sim.10064

ElSayed NA, Aleppo G, Aroda VR, et al.; American Diabetes Association. 6. Glycemic Targets: Standards of Care in Diabetes-2023. *Diabetes Care*. 2023:46(suppl 1):S97-S110. https://doi.org/10.2337/dc23-S006

[EMA] European Medicines Agency. Guidance document for the content of the <Co->Rapporteur day 80 critical assessment report. Clinical aspects EMA/269176/2014. Published 2014. Accessed May 3, 2023. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/day-80-assessment-report-clinical-guidance_en.pdf

EuroQol Research Foundation. EQ-5D-5L User Guide, Version 3.0. Updated September 2019. Accessed July 29, 2022. https://euroqol.org/publications/user-guides

[FDA] United States Food and Drug Administration. Adjusting for covariates in randomized clinical trials for drugs and biological products: guidance document: guidance for industry. May 2023. Accessed July 24, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products.

https://www.fda.gov/drugs/regulatory-science-action/impact-story-using-innovative-statistical-approaches-provide-most-reliable-treatment-outcomes

[ICH] International Council for Harmonisation. ICH harmonised guideline. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for

clinical trials. E9(R1). November 20, 2019. Accessed November 07, 2022. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

[JASSO] Japan Society for the Study of Obesity. Guideline for the management of obesity disease 2022. Tokyo: Life Science Publishing. 2022. (in Japanese)

Kolotkin RL, Ervin CM, Meincke HH, et al. Development of a clinical trials version of the Impact of Weight on Quality of Life-Lite questionnaire (IWQOL-Lite Clinical Trials Version): results from two qualitative studies. *Clin Obes*. 2017;7(5):290-299. https://doi.org/10.1111/cob.12197

Kolotkin RL, Williams VSL, Ervin CM, et al. Validation of a new measure of quality of life in obesity trials: Impact of Weight on Quality of Life-Lite Clinical Trials Version. *Clin Obes*. 2019;9(3):e12310. https://doi.org/10.1111/cob.12310

Ma C, Shen X, Qu Y, Du Y. Analysis of an incomplete binary outcome dichotomized from an underlying continuous variable in clinical trials. *Pharm Stat*. 2022;21(5):907-918. https://doi.org/10.1002/pst.2204

Maruish ME, editor. *User's Manual for the SF-36v2 Health Survey*. 3rd ed: QualityMetric Inc; 2011.

PHUSE. Analysis and displays associated with adverse events: focus on adverse events in Phase 2-4 clinical trials and integrated summary documents. Published February 03, 2017. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+and+Displays+Associated+with+Adverse+Events+Focus+on+Adverse+Events+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

PHUSE. Analysis and displays associated with demographics, disposition, and medications. Published March 02, 2018. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Demographics,+Disposition+and+Medication+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary+Documents.pdf

PHUSE. Analyses and displays associated with laboratory analyte measurements in Phase 2-4 clinical trials and integrated submission documents – update to recommendations. Published August 19, 2022. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Safety+Analytics/WP068.pdf

PHUSE. Analyses and displays associated with measures of central tendency – focus on vital sign, electrocardiogram, and laboratory analyte measurements in Phase 2-4 clinical trials and integrated submission documents. Published October 10, 2013. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Measures+of+Central+Tendency-+Focus+on+Vital+Sign,+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Submissions.pdf

PHUSE. Analyses and displays associated with outliers or shifts from normal to abnormal: focus on vital signs, electrocardiogram, and laboratory analyte measurements in Phase 2-4 clinical trials and integrated summary documents. Published September 10, 2015. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Outliers+or+Shifts+from+Normal+To+Abnormal+Focus+on+Vital+Signes+%26+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

Rothmann M. Use of Bayesian hierarchical models in the presentation of subgroup analyses. Slides presented at: DIA Bayesian Scientific Working Group Webinar Series; November 19, 2021; Washington DC.

Röver C, Bender R, Dias S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res Synth Methods*. 202112(4):448-474. https://doi.org/10.1002/jrsm.1475.

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc.; 1987.

Wang B, Du Y. Improving the mixed model for repeated measures to robustly increase precision in randomized trials [published online November 29, 2023]. *Int J Biostat*. https://doi.org/10.1515/ijb-2022-0101

Ye T, Bannick M, Yi Y, Shao J. Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. *Stat Theory Relat Fields.* 2023;7(2):159-163. https://doi.org/10.1080/24754269.2023.2205802

Ye T, Shao J, Yi Y, Zhao Q. Toward better practice of covariate adjustment in analyzing randomized clinical trials. *J Am Stat Assoc*. 2022;118(544):2370-2382. https://doi.org/10.1080/01621459.2022.2049278

Signature Page for VV-CLIN-137368 v2.0

Approval	PPD 05-Aug-2025 14:07:28 GMT+0000

Signature Page for VV-CLIN-137368 v2.0

Approved on 05 Aug 2025 GMT